AN INVESTIGATION OF THE STRUCTURES OF SOME CRYSTALLINE PEPTIDES AND THE SYNTHETIC POLYPEPTIDE POLY-γ-METHYL-L-GLUTAMATE

Thesis by

Harry L. Yakel, Jr.

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

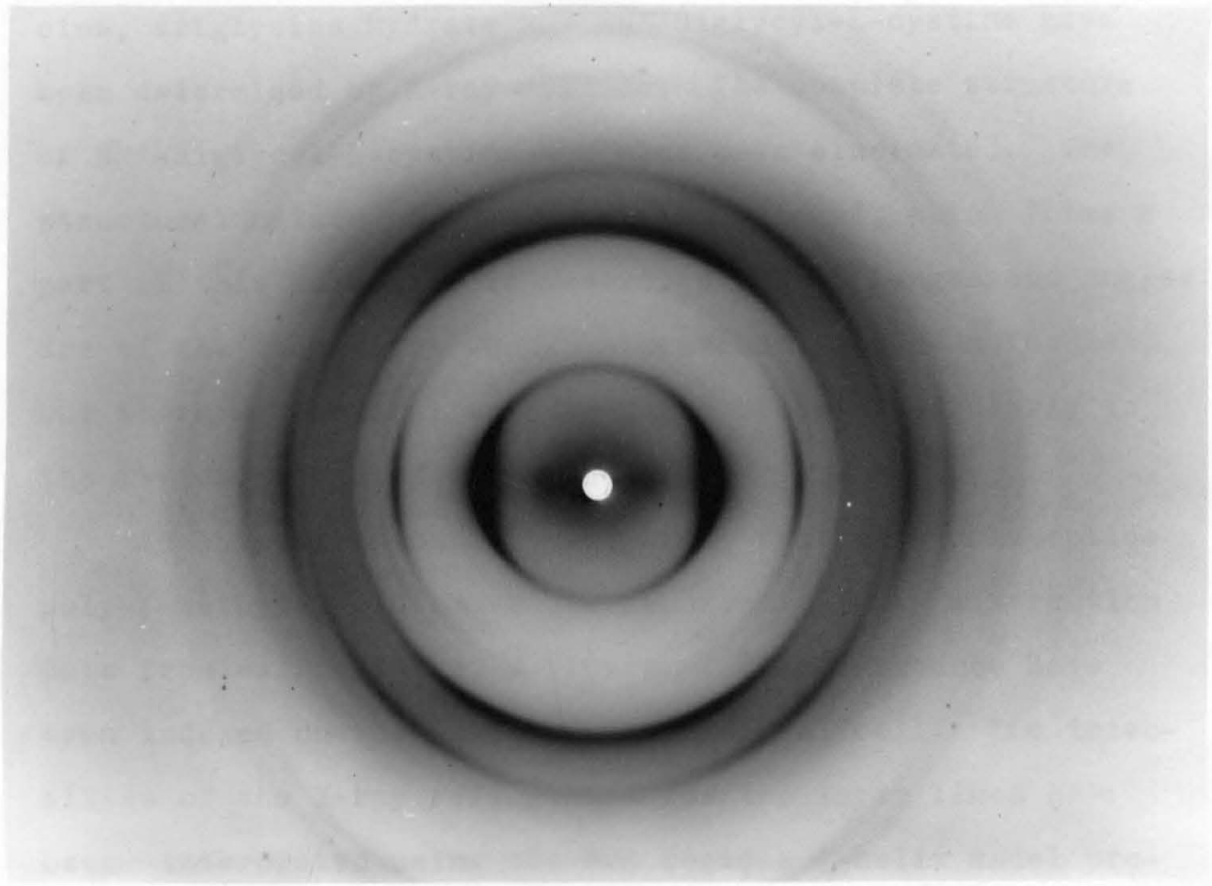
Pasadena, California

1952

ACKNOWLEDGEMENT

The author is particularly indebted to Dr. E.W. Hughes, Dr. Linus Pauling and Dr. R.B. Corey who were so liberal with their advice and comments during the course of this work. Thanks are also due to Drs. Niemann, Schomaker, and Badger for their assistance in some of the problems connected with the investigation of poly-γ-methyl-L-glutamate.

The author also wishes to thank William Chaplin of the Digital Computing Laboratory for his aid in setting up the numerical calculations involved in the determination and refinement of the structure of NN'-diglycyl-L-cystine.



X-ray photograph of oriented poly-γ-methyl-L-glutamate (α form)

ABSTRACT

The unit cell dimensions and space groups of α-triglycine, triglycine hydrate and NN'-diglycyl-L-cystine have been determined by X-ray methods. The complete structure of NN'-diglycyl-L-cystine has also been elucidated. The structural nature of the "disulfide bridge", which forms a part of this molecule, is described. Bond lengths and angles are of the usual size found in previous peptide molecules, but there are several weak hydrogen bonds particularly in the amide groups of the molecule.

An investigation of the fibrous synthetic polypeptide poly-γ-methyl-L-glutamate is reported. X-ray diffraction data from well oriented α fibers of this substance have been indexed on the basis of a hexagonal cell. The intensities of the X-ray reflections for two layer lines have been interpreted using the 3.6 residue α-helix model proposed by Pauling and Corey. A resume of the data appears to support this model for the polypeptide chain in poly-γ-methyl-L-glutamate.

TABLE OF CONTENTS

PART	TITLE	PAGE
I.	INTRODUCTION	1
	A) Investigations of the structures of crystalline amino acids and peptides	2
	B) Research on proteins and synthetic polypeptides	4
	1. Classification of proteins	4
	2. A general background of studies on fibrous proteins	6
	3. Proposed models for the folded polypeptide chain	8
	C) The purpose of this thesis	12
II.	THE DETERMINATION OF THE UNIT CELL DIMENSIONS AND SPACE GROUPS OF SEVERAL SIMPLE CRYSTALLINE PEPTIDES	14
	A) α-Triglycine	14
	B) Triglycine hydrate	20
	C) NN'-Diglycyl-L-cystine	25
III.	THE COMPLETE INVESTIGATION OF THE STRUCTURE OF NN'-DIGLYCYL-L-CYSTINE	30
	A) The determination of the trial structure	30
	1. Interpretation of the hOL Patterson projection	30
	2. The preliminary refinement of the x and z parameters	36
	B) The completion of the structure analysis and final refinement of parameters	52
	1. An estimation of y parameters aided by a calculation of several line Pattersons and Patterson sections	52
	2. The refinement of the structure using three-dimensional least squares methods	61
	C) A discussion of the results of this investigation	83

PART	TITLE	PAGE
IV.	A STUDY OF THE STRUCTURE OF POLY-γ-METHYL-L-GLUTAMATE.	94
	A) Preparation of oriented fibers of the polypeptide	94
	B) A careful measurement of spacings on X-ray photographs of oriented poly-γ-methyl-L-glutamate	.100
	1. Measurements of reflections observed using a camera of 3 cm. radius. . . .	.100
	2. Measurements on photographs taken in a camera of 10 cm. radius	.102
	C) Calculations of X-ray diffraction intensities for the α-helix model of poly-γ-methyl-L-glutamate	.105
	D) A critical evaluation of the results. . . .	.120
V.	APPENDICES	
	I. The attempted application of Banerjee's equation to the solution of the α-triglycine structure	.126
	II. A summary of the punched card techniques utilized in the refinement of the NN'-diglycyl-L-cystine structure	.129
VI.	REFERENCES	.132
VII.	PROPOSITIONS	.135

I. INTRODUCTION

The great interest in the field of protein structure has been given marked impetus by the intensive application of X-ray techniques to its manifold problems. Many other physico-chemical methods have been applied to these same problems, but it seems apparent that X-ray methods have in the past yielded more fruitful results than any other, and that the final solution of the general question of protein structure will eventually be provided by systematic X-ray investigations.

It has been firmly established by means of chemical analyses on protein hydrolysates that most, if not all, proteins contain long chains of amino acids connected by amide linkages -- so-called polypeptide chains. While these chemical investigations do serve to indicate the order of amino acid residues in the polypeptide chains, they do not shed any light on their structural properties. The elucidation and explanation of these properties must wait on the interpretation of microscopic, spectrographic, and especially X-ray data.

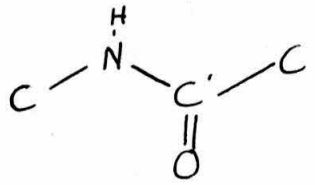
Historically, research into the structure of proteins and polypeptides has followed two distinct paths. On the one hand, considerable energy has been devoted to the determination of the structural features of the simple amino acids and peptides on the hypothesis that a direct correspondence between the structural characteristics of the pro-

tein molecule and those of the building stones from which it is constructed should exist. The second line of investigation has concerned itself with direct observations on natural proteins and synthetic polypeptides. It is difficult to say which of these two approaches to the problem has had the greater success to date: the ideal toward which each path is aimed is a structural model which will explain the results of both.

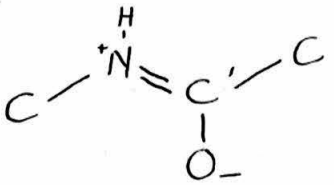
A) Investigations of the structures of crystalline amino acids and peptides.

As examples of work along the first of these paths the determinations of the structures of the amino acids glycine (1), alanine (2), L-threonine(3), and serine (4), from X-ray data, may be cited. These researches beside providing reliable values for bond lengths and angles have emphasized the important role which hydrogen bonds play in the configurations of these simple molecules and, presumably, in the configurations to be expected in protein molecules.

Another very important structural characteristic which should exist in the polypeptide chains of proteins became apparent both through theoretical considerations by Pauling and through actual determinations of peptide structures. This is the coplanarity of the atoms involved in the amide link connecting amino acid residues. The necessity for coplanarity arises from the resonance of forms I and II in the imide group. It is well known that resonance of this



I



II

keto-enol type requires a dihedral angle of approximately zero degrees between the plane of the CNH atoms and the plane of the NC'OC atoms for the most effective interaction of the pi orbitals of the nitrogen and carbonyl carbon atoms. This restriction, deduced from the resonance theory of Pauling, was found to hold within the limits of experimental accuracy when the results of the crystal structure analyses of diketopiperazine(5), N-acetyl glycine (6), and α and β-diglycine (7,8) were examined. In further agreement with resonance principles, it was also possible to correlate the observed C-N bond lengths with the distances predicted by Pauling's equation (9) connecting bond character and length.

Although the principle of coplanarity of the imide group has been shown to be quite general in the structures of simple peptides, it has not been universally accepted as applicable to proteins and polypeptides. As recently as 1951, Ambrose and Elliott (10) refer to the question in the following words: "It remains to be shown whether resonance in polypeptides and proteins effectively limits the rotation about the C-N bond, and we regard the matter as by no means settled." It should be noted, however, that one of

the more successful models of the polypeptide chain proposed to date, the α helix, incorporates this structural feature.

In passing, it must be pointed out that complete structural examinations of even the simple amino acids and peptides, using X-ray data, involve computational difficulties which would prove practically insurmountable were it not for the development of high speed automatic calculating machines. With the advent of punched card methods, however, structures involving upward of twenty five positional parameters have been solved in a moderate length of time.

An excellent review of the results gleaned from various crystal structure determinations on molecules containing peptide, or peptide-like, configurations is to be found in an article by R.B. Corey and J. Donohue (11) appearing in the Journal of the American Chemical Society.

B) Research on proteins and synthetic polypeptides

1. Classification of proteins

Turning now to the second path of research on the problem of protein structure, investigations on proteins and synthetic polypeptides, a further delineation is possible on the basis of the type of protein selected for study. Although many schemes of classification of proteins have been proposed, the most useful from the standpoint of the structural chemist is that which differentiates the general group into globular and fibrous proteins. (While they themselves are not proteins, the many similarities between synthetic

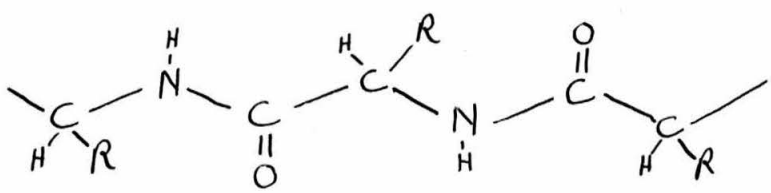
polypeptides and natural fibrous proteins justifies the inclusion of the former substances in the class of fibrous proteins.) The names are descriptive of the over-all structural nature of the respective proteins, the globular protein molecules being approximately spherical in shape in contrast with the chain-like extended molecules classified as fibrous proteins.

From the difference in shape alone, one might predict that the globular proteins would be readily crystallized whereas the fibrous proteins would tend to give pseudo-crystals ordered in the fiber direction but greatly disordered in directions perpendicular to the fiber. That this is found to be the case does not mean that X-ray investigations of fibrous proteins and synthetic polypeptides have been less productive of results than single crystal work on globular proteins. Indeed, while Bernal (12), Perutz (13), and Harker (14) have contributed, and are at present contributing, much to the understanding of protein structure through their work on globular molecules such as hemoglobin, some of the most interesting developments in this field have emerged from studies of fibrous proteins.

Because a part of the research reported in this thesis deals with an investigation of a fibrous poly-peptide, it is perhaps pertinent to discuss here the history of work on this class of proteins without giving in any detail a similar description of globular protein studies.

2. A general background of research on fibrous proteins

It has long been recognized that many naturally occurring materials, for example muscle, hair, silk, gelatin, etc., are proteins of the fibrous variety. Tests on the products of acid, base, and enzyme hydrolyses have shown the presence of a great many kinds of amino acids in these proteins. This immediately introduces the problem of uncertain chemical composition and order of amino acid residues into any attempted X-ray examination of these substances. Still a great number of empirical data and some illuminating deductions were derived from early X-ray work in this field. In particular, Astbury and Woods (15) found that the polypeptide chains in these fibers could be extended by mechanical stretching or by treatment with formic acid. Analysis of the X-ray patterns given by the extended fibers demonstrated that the chains existed in the nearly fully stretched, zig-zag configuration shown below. This configuration has been termed the β form of the polypeptide chain.



Since the β form is obtained by drastically stretching the natural protein, it follows that previous to the stretching process the polypeptide chains were coiled in

some manner. The problem of precisely how the chains are folded in the so-called α form of fibrous proteins has been the subject of intensive research in recent years.

Before discussing some of the proposed models of the coiled polypeptide chain, a description of the experimental difficulties inherent in X-ray work on fibrous proteins should be given. The rather random orientation to be expected in untreated fibers produces a diffraction pattern containing fairly long and diffuse arcs in place of the sharp spots usually associated with diffraction from single crystals. The character of these reflections makes precise measurements on the films virtually impossible, and these inaccuracies in measurements are reflected in the large probable errors in the lattice spacings derived from them. Much of the controversy over the validity of the various models proposed for the coiled chains has occurred only because accurate values for the interplanar spacings of these structures have not been available. It has been found that slight stretching of the fibers tends to sharpen the reflections considerably, but much of the ambiguity still remains.

It has also been observed that the diffraction patterns given by suitably treated fibers of synthetic polypeptides often show more orientation than is obtainable with natural proteins (16). These synthetic substances, prepared by polymerization of a simple amino acid, possess the advantage of known composition, but lack the uniformity of chain length found in the natural proteins. Their marked similarity to

the naturally occurring polypeptides (they both exist in α and β forms, and both have many common details in their diffraction patterns) have made these compounds the subjects of numerous investigations.

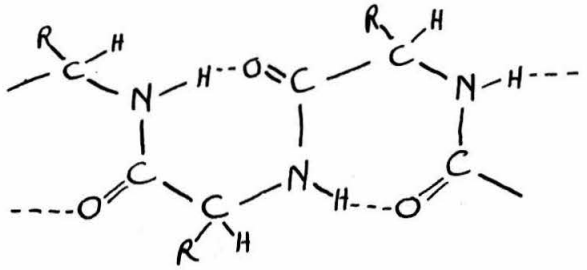
3. Proposed models for the folded polypeptide chain

One of the earliest models of the folded chain in the α form of polypeptides was proposed simultaneously by Huggins (17), and Astbury and Bell (18). This model involves a rather flat loop of three amino acid residues per repeat unit in the fiber direction, accounting for the name α_{III} given to this configuration. The repeat in the fiber direction could be calculated from bond distances and angles current at the time as 5.15A. The actual diffraction data on fibrous proteins and synthetic polypeptides were then so meager, however, that no critical evaluation could be made.

Bragg, Kendrew, and Perutz (19), in 1950, published a paper discussing several possible configurations of a coiled polypeptide chain, including a helical model which contained four amino acid residues per turn of the helix. Working at the time on crystalline hemoglobin and myoglobin, which globular proteins are also believed to contain folded chains, these authors compared their various models with Patterson projections of the protein structures and concluded that the helical model must be eliminated in favor of Astbury's α_{III} configuration.

Ambrose and Hanby (20), studying the infra-red absorp-

tion spectra of oriented samples of synthetic polypeptides, discussed another model for the folded peptide chains in these substances to account for the observed dichroism of the N-H and C-O stretching bands. This model, called the α_{II} configuration, consists of a practically flat loop of two amino acid residues in a repeat of about 5.14A. (See sketch below.)



Bamford, Hanby, and Happey (21), working with the X-ray patterns from these same synthetic polypeptides, also were able to find evidence in support of the α_{II} model. In their publication, the authors state that the reflections from poly-γ-methyl-L-glutamate fibers can be accounted for, with but two exceptions, by an orthorhombic cell with parameters a = 10.35A., b = 5.98A., and c = 5.50A. The spacings which do not fit this cell are ascribed to a second phase coexistent with the α phase. Similar analysis of the pattern given by poly-γ-benzyl-L-glutamate suggest a pseudo-monoclinic cell according to Bamford. Again in this pattern they find reflections which must be attributed to a second phase.

It has been pointed out by Pauling and Corey (22) that the various structures of the coiled polypeptide chain proposed thus far are not entirely in agreement with the results obtained from crystalline peptides and amino acids.

For example, in the α_{III} model only one out of every three amide groups forms a hydrogen bond, while it has been observed that numerous strong hydrogen bonds are invariably found in simple peptide structures. It has also been shown that in order to maintain reasonable hydrogen bond distances in the α_{II} model, the coplanarity of the amide link must be destroyed. These considerations, together with several other objections regarding the $\alpha \rightarrow \beta$ transformation and the dichroic ratio of the N-H and C-O stretching bands, prompted Pauling to examine additional models for the folded peptide chain which would not possess these anomalies.

The models evolved by Pauling and Corey (22) are helical in shape, but differ from the helix proposed by Bragg, Kendrew, and Perutz in that they do not possess an integral number of amino acid residues per turn of the helix. The different helices are named according to the number of residues per turn; for example the 3.7 helix. In a series of papers published in the Proceedings of the National Academy of Sciences (22), Pauling and Corey demonstrate the agreement between the results predictable on the basis of the helical models and the observed data on synthetic polypeptides. (It should be noted that although similar helical structures have been suggested for some naturally occurring proteins, up to the present time they have found their greatest success as models for synthetic polypeptides.)

In particular, Pauling and Corey are able to interpret the poly-γ-methyl-L-glutamate data of Bamford, et al., by

means of a hexagonal cell (a = 11.7A., c = 27.5A.) in which the repeat in the fiber direction is five times the repeat proposed by Bamford. Pauling assumes a molecular structure having a translation in the chain direction of 1.5A. per residue, giving a total of about 18 residues and 5 turns per repeat distance, or 3.7 residues per turn. The data on poly-γ-benzyl-L-glutamate are also discussed by these authors in the light of the α helices. For this compound they find that a helix containing 11 residues in 3 turns, another variation of the 3.7 helix, explains Bamford's data satisfactorily.

It is interesting to note that the second phase which Bamford assumes present in the fibers in order to account for all the observed reflections is not necessary in Pauling and Corey's picture. The anomalous arcs can be indexed without difficulty on the basis of the 11.7 by 27.5A. hexagonal cell.

After the appearance of these papers, Perutz (23) suggested that if the helical models were correct the translational rise per residue in the fiber direction should produce a rather strong reflection corresponding to that distance on the meridian of oriented fiber photographs. Immediate investigation of poly-γ-methyl and poly-γ-benzyl-L-glutamate confirmed that these reflections were present and gave considerable support to the Pauling-Corey models. Since that time it has been found that spacings similar in magnitude and position to the Perutz reflection occur in a great many natural fibrous and globular proteins. This suggests that the helical configurations, or something closely allied to them,

may exist in these substances too.

Although this necessarily brief account of the general problem of protein structure is far from complete, most of the topics pertinent to the research reported here have been covered. At various places in the body of the thesis references to previous work may be made in more detail, but it was not deemed advisable to encumber this introduction with either lengthy descriptions of relatively minor points or critiques which would be more aptly included in a discussion of results.

C) The purpose of this thesis

The field of protein structure having been rapidly surveyed and a few of its more important questions dealt with in some detail, it is appropriate that a statement of the purpose of this thesis be made.

First, the author proposes to present a description of some rather elementary work on three crystalline peptides, α-triglycine, a triglycine hydrate, and NN'-diglycyl-L-cystine. This will be followed by a detailed account of the complete determination of the crystal structure of NN'-diglycyl-L-cystine.

Secondly, the author proposes to describe some recent work on the structure of the synthetic polypeptide poly-γ-methyl-L-glutamate. This investigation deals particularly with an attempt to obtain accurate interplanar spacing measurements and to obtain further evidence for or against the

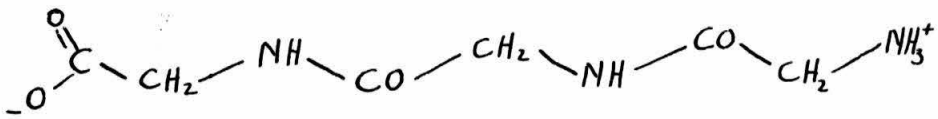
α helix model of Pauling and Corey.

Following both the section devoted to crystalline peptides and that dealing with synthetic polypeptides a critical discussion of the significance of the results as applied to the question of protein structure will be presented.

II. THE DETERMINATION OF THE UNIT CELL DIMENSIONS AND SPACE GROUPS OF SEVERAL CRYSTALLINE PEPTIDES

A) α-Triclycine

In an attempt to add to the general fund of data on crystalline peptides, it was deemed advisable to begin an investigation of a tripeptide. Sketchy unit cell parameters and space groups for such compounds have been reported, but no serious structural analysis has been undertaken to date. Owing to its chemical simplicity triglycine (glycylglycylglycine) was a likely choice for this study. In the usual representation of organic chemistry, triglycine has the following formula:



A small sample of crystalline triglycine prepared by Dr. J. Greenstein was made available for this investigation by the Harvard Medical School. The material was dissolved in water and recrystallized by the slow addition of n-propanol. Fine needles began separating from the solution at a composition of about 80% water: 20% alcohol.

Microscopic examination of the crystals showed them to be in reality very thin laths with one set of rather well developed faces parallel to the needle axis. No remarkable cleavage properties were observed, although the crystals did tend to fracture along a plane perpendicular to their long axes.

Inspection with crossed Nicols showed the crystals to be highly birefringent with extinctions parallel and perpendicular to the needle axis. By switching to a high power objective and removing the eyepiece, the characteristic biaxial interference figure was observed. The over-all conclusion reached by means of this microscopic analysis was that the crystals were probably either monoclinic or orthorhombic with one very short and two moderately long unit cell axes.

A study of the literature revealed that several forms of crystalline triglycine had already been reported. One of these, an orthorhombic dihydrate, was investigated by Bernal (24), while two anhydrous modifications, called α and β-triglycine, were briefly studied by Lenel (25). Although far from conclusive, the fact that heating produced no characteristic efflorescence indicated that the crystals under investigation were anhydrous.

One of the larger crystals from the preparation was mounted with the needle axis vertical and examined goniometrically. Aside from signals from the flat lath faces, no recognizable reflections were obtained. An oscillation photograph, with copper K radiation, was taken and the repeat distance in the needle direction determined as approximately 4.35kX. Comparison with unit cell dimensions for triglycine dihydrate and α and β-triglycine showed that the closest agreement was obtained with α-triglycine, for which Lenel reports a b axis of 4.3A.

A series of oscillation photographs was taken and a general indexing attempted with a reciprocal lattice net constructed with Lenel's lattice parameters for α-triclycine. It was found that the entire zero layer could be indexed satisfactorily, but several strong reflections on the first layer line could not be accounted for. By doubling the c axis given by Lenel, these reflections could be indexed. Since it was extremely difficult to mount a crystal with the c axis vertical, final confirmation of the doubled c axis was deferred until a complete set of Weissenberg photographs was available.

Accurate lattice parameters were next obtained from double oscillation pictures taken so as to produce a particular reflection on both sides of the film. Precise measurement of the position of the 600 reflection gave a*, while c* was determined from the 008 reflection (based on Lenel's cell), and β* from the 606 reflection. Referring back to the real lattice, the following parameters were obtained:

Cell Dimensions for α-Triglycine

Lenel	This Investigation
a = 8.53kX	a = 8.53kX ± 0.01kX
c = 11.4kX	c = 11.48kX or 22.96kX ± 0.01kX
β = 105°30'	β = 105°36' ± 10'
b = 4.3kX	b = 4.35kX ± 0.02kX

Lenel's parameters are included for purposes of compari-

son. It is seen that the agreement is very good except for the c axis, which was believed to be 22.96kX instead of 11.4kX as given in the earlier work. The b distance reported for this investigation was obtained by averaging a number of layer line spacing measurements and is undoubtedly not as accurate as the a, c, or β parameters.

A set of hOL, h1L, h2L, h3L, OkL, and 1kL Weissenberg photographs were taken i) to obtain intensity data, ii) to confirm the doubled c axis, and iii) to fix the space group. These pictures clearly demonstrated that the c axis was indeed about twice as long as proposed by Lenel. The only systematic extinctions observed were in the hOL zone, all reflections for which L was odd being absent, and in the OkO zone, all reflections with k odd being missing. The extremely short b axis gave rise to only four observable rows on the OkL Weissenberg and, on these lines, it was seen that the 010, 020, and 030 reflections were absent, with a weak 040 being the only order present. Thus the assignment of a systematic extinction to this zone might be viewed with some scepticism. Assuming these two systematic extinctions, however, the space group $P2_1/c$ was unambiguously deduced.

The crystal density was determined by a flotation method as 1.56 ± 0.03 gm./cm^3. This density, in conjunction with the unit cell dimensions given above and the known molecular weight of the compound, indicated that there were four molecules per cell. As is to be expected in a case involving an unsymmetrical molecule, the atoms all lie in general positions.

Measurements of reflection intensities were all made by visual comparison of the spot in question with a calibrated intensity strip. This strip was prepared with an hOL reflection from α-triglycine so that the shapes of the spots being compared were essentially the same. Only hOL zone intensities were recorded at this time since the preliminary work on the structure would obviously be done with projections in that zone. Multiple films were used for each Weissenberg and, with the aid of a suitable film factor, were of assistance in the estimation of intensities.

The measured intensities were multiplied by Lorentz and polarization factors obtained graphically by the method of Lu (26) to give corrected intensities. Absolute intensities and an approximate temperature factor were then obtained for the hOL zone according to the method developed by A.J.C. Wilson (27).

As has been suggested above, the short b axis made it very probable that as much information could be derived from two-dimensional syntheses as from three dimensional work. Therefore the corrected absolute hOL intensities were used to compute an hOL Patterson projection. This unsharpened projection is shown in Fig. I. The presence of negative areas is probably due to a slight error in the scale factor.

The row of peaks nearly parallel to w and at u = 0 is in many ways similar to the row of peaks observed in nearly the same position in the Patterson projection of the β-diglycine structure. This strongly suggested that, as in the

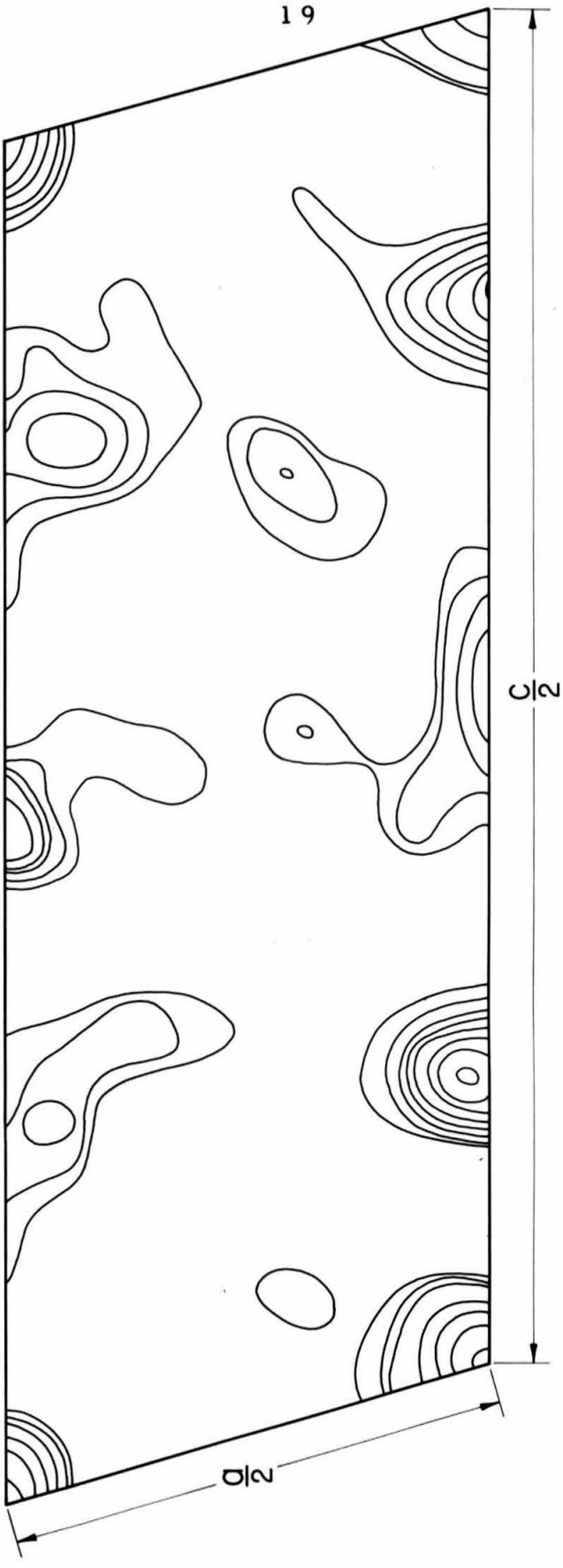


Fig. I. hOl PATTERSON PROJECTION FOR α-TRIGLYCINE

dipeptide, the tripeptide chain lies parallel to c. The long c axis bears out this hypothesis as does the unusually high intensity of the 200 reflection. The general picture of the structure as derived from the Patterson, then, is that the triglycine molecules in the fully extended form are lying nearly parallel to c and along a line at roughly x = 1/4.

It should be noted that various trial structures have been proposed in accordance with these criteria, but none seems to give calculated intensities in agreement with the observed. A discussion of the attempted application of Banerjee's equation to this problem will be given in an appendix to this thesis.

B) Triglycine hydrate

Previous to the successful crystallization of α-triglycine, a saturated solution of triglycine in water had been allowed to evaporate slowly, with a small amount of n-propanol added to promote crystal formation. The crystals obtained from this preparation were very similar in shape to the α-triglycine needles encountered later. By studying them under the polarizing microscope the crystals were seen to be strongly birefringent with a biaxial interference figure.

Slight heating of the crystals caused them to effloresce, from which it was concluded that the crystals were hydrated. As has been mentioned in the last section, the literature indicates that three forms of triglycine had been found and

studied prior to this work. Since only one of these forms was a hydrate (reported by Bernal (24)), it was believed that the modification found here might well be identical with this orthorhombic dihydrate.

A suitable crystal being mounted with the needle axis vertical, rotation and oscillation pictures with copper K radiation were taken. The needle axis repeat as obtained from the layer line spacings on these photographs was about 4.6A., in fair agreement with the 4.7A. repeat found by Bernal for the dihydrate. However, the rotation pictures failed to show the symmetry which would be expected for an orthorhombic crystal. To decide the point a zero layer Weissenberg about the needle axis was made. This picture did not show any prominent lines of reflections which could be ascribed to axes 90° apart, but instead indicated that the a and c axes were about 112° apart. The monoclinic symmetry was confirmed by a set of first layer Weissenbergs.

The determination of accurate lattice parameters was accomplished essentially in the same way as in the case of α-triglycine. The results are given below together with the reported values for the other three forms of triglycine. It is clear that the hydrate found here is a new modification of triglycine.

Compound	a	b	c	β
α-Triglycine (Lenel)	8.53kX	4.3kX	11.4kX	105°30'
β-Triglycine	14.6kX	4.79kX	11.67kX	105°30'
Triglycine Dihydrate	22.0kX	9.8kX	4.7kX	---
Triglycine x-hydrate	16.02kX	4.63kX	24.99kX	112°43'

The density of the crystal was also determined using the flotation method employed for α-triglycine. The value obtained, 1.52 ± 0.02 gm/cm^3, coupled with the unit cell dimensions given above, showed that there were eight molecules per cell. The molecular weight used in this calculation corresponded to the formula $C_6N_3O_4H_{11} \cdot 1/2\ H_2O$ with one water of hydration for every two peptide molecules. Although the experimental density was by no means accurate enough to clearly establish this degree of hydration, it was noted that the best agreement was obtained for the hemihydrate. A monohydrate would correspond to a calculated X-ray density of 1.56 gm/cm^3 which, while not positively excluded, is near the upper limit for the observed density. A vacuum desiccation of a small batch of these crystals was attempted, but here the quantity of material was too small to permit a definite degree of hydration to be calculated.

hOL and hlL Weissenbergs showed the extinctions characteristic of A face centering and an a glide plane. The space group could not unambiguously be assigned on this basis since either the Aa or A2/a space groups satisfied both extinctions.

However, as the general positions of the Aa space group are fourfold while those of the A2/a space group are eightfold, the latter was chosen as a starting hypothesis. No piezo- or pyro-electric measurements were made, so the Aa space group still remains a possibility. There would then be two independent sets of triglycine molecules occupying general positions.

Intensities of the hOL reflections were visually estimated using an intensity strip. Lorentz and polarization factors calculated by Lu's graphical method were applied and an hOL Patterson projection synthesized. As in the case of α-triglycine, it was hoped that the projection down the short 4.63kX axis would give most of the information needed for the deduction of a trial structure. The unsharpened Patterson, on an arbitrary scale, is shown in Fig. II.

The row of peaks lying roughly parallel to the $40\bar{6}$ plane in the projection suggests that the triglycine molecules may be oriented in this direction. This is in agreement with the result obtained from the examination of the birefringence of these crystals which showed that the positive orientation of the γ axis of the indicatrix lay in the $40\bar{6}$ direction.

No further interpretation of this Patterson projection has been made to date. It is very likely that an exhaustive analysis could be made to yield a satisfactory trial structure, but this was not attempted as the preliminary work on NN'-diglycyl-L-cystine showed greater promise of success.

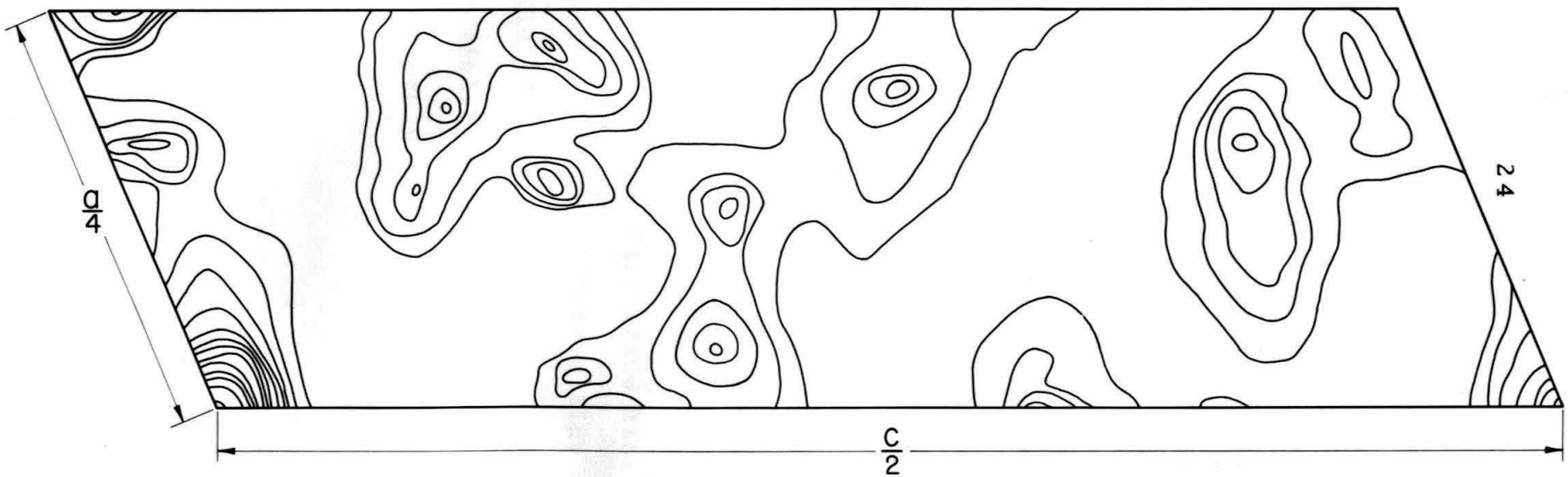
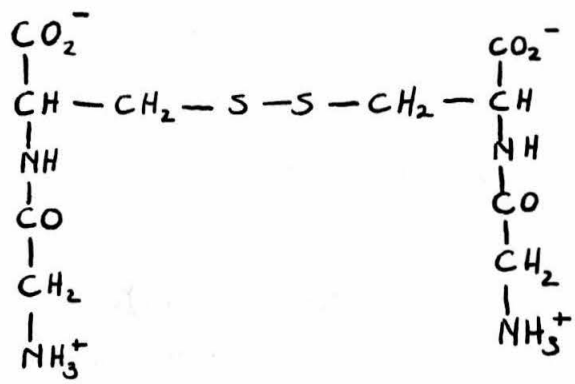


Fig. II. hOl PATTERSON PROJECTION FOR TRIGLYCINE HYDRATE

C) NN'-Diglycyl-L-cystine

As has been mentioned in the introduction to this thesis, the structures of various amino acids and peptides shed considerable light on the nature of the hydrogen bonding which can be expected between polypeptide chains, and on the probable interatomic distances and configurations which may occur in these chains. Previous to this research, however, no structural investigation of a molecule containing the so-called "disulfide link" has been reported. This link, which constitutes one of the more important ways in which the peptide chains of proteins are believed to be held together, consists of a $-CH_2-S-S-CH_2-$ bridge connected to a polypeptide chain at either end. The fact that the peptide chain at one end of the bridge may be identical with that at the other end, a doubling back of the chain having taken place, also indicates that the "disulfide link" may play an important role in the folding of polypeptide chains.

A simple peptide which could contain such a configuration is NN'-diglycyl-L-cystine, first prepared by Greenstein (28). The molecule has the following formula:



A sample of this compound being kindly provided by Dr. B. Low of the Harvard Medical School, an investigation was begun in an effort to elucidate the structural characteristics of the "disulfide link".

The powdered sample of NN′-diglycyl-L-cystine was dissolved in water and recrystallized by the dropwise addition of n-propanol. The crystals were similar in shape to the d-triglycine crystals, separating from solution in the form of flat needles with one pair of well developed faces parallel to the needle axis. They were strongly birefringent, giving extinctions parallel and perpendicular to the needle axis, and also showed a biaxial interference figure. These facts, although not conclusive, indicated monoclinic symmetry.

A crystal was mounted with the needle axis vertical and an oscillation picture with copper K radiation taken. This showed the repeat along the needle direction to be about 4.8Å. From a zero layer Weissenberg it was noted that, selecting an arbitrary a axis, two c axes could be chosen, one at an angle of 100° from a and the other at an angle of about 124°.

A first layer Weissenberg confirmed the monoclinic symmetry of the crystal and also made possible a unique determination of the space group. With the c axis at $\beta = 100^\circ$ it was evident that the observed extinctions required an F2 space group, whereas with the alternate axis at 124°, an A2 space group resulted. Because of the smaller number of mole-

cules per cell in the A2 space group, the axis at 124° was selected as the c axis of the cell.

In order to obtain accurate unit cell dimensions, oscillation photographs using a 5 cm. camera were employed. The same double oscillation technique used to procure parameters for the α-triglycine and triglycine hydrate structures was utilized. Oscillations around the needle axis yielded the following results for a, c, and the angle β:

$$a = 12.26A.;\ c = 17.17A.;\ \beta = 124°24'.$$

Moreover a careful measurement of the layer line spacings on these same photographs gave the value of 4.84A. for b. The probable errors in these parameters are of the order of ± 0.01kX.

In his paper reporting the preparation of NN'-diglycyl-L-cystine (28), Greenstein states that the crystals are hydrated, there being one water molecule for every diglycyl-cystine molecule. The empirical formula $C_{10}H_{18}N_4S_2O_6 \cdot H_2O$ corresponds to a molecular weight of 370. A rough density determination using a flotation method gave a value of 1.55 ± 0.04 gm/cm^3. With the cell dimensions given above, a simple calculation showed that the number of molecules per cell must be two. As the general positions of the A2 space group are fourfold, this result indicated that the diglycyl-cystine molecules were lying in special positions, probably across twofold axes since the peptide molecule itself would be expected to contain this symmetry element. Assuming the degree of hydration reported by Greenstein to be correct,

the water molecules must also occupy special positions, presumably on the twofold screw axes present in the structure. Such an arrangement would suggest some sort of chain of water molecules in the b direction interconnected by means of hydrogen bonds, but with a very short O - H -- O distance of only 2.42A. and a rather odd disposition of H - O --- H bond angles. Because of these discrepancies with previously recorded results, Greenstein's degree of hydration was regarded with some suspicion, and, as will be shown later, was actually found to be incorrect.

As has been mentioned, hOL and h1L Weissenbergs were taken to fix the crystal symmetry and to aid in the selection of cell axes. h2L, h3L, and OkL Weissenbergs were also taken, using copper K radiation for all. Multiple films were employed to assist in estimation of intensities. As in the case of the triglycine compounds, all intensities were estimated visually from a calibrated intensity strip. Lorentz and polarization factors were read from graphs of the factors as functions of sin θ and the inclination angle μ. No correction for absorption was made since the crystal selected for examination was very small in crossection. The scale factor and an approximate temperature factor for the corrected hOL zone reflections was determined by the method of A.J.C. Wilson (27). No attempt was made to put the other reflections on an absolute scale at this stage.

A description of the computation and interpretation of the Patterson syntheses for NN'-diglycyl-L-cystine will not

be given here since the next section will present a complete structural analysis and refinement for this compound.

III. THE COMPLETE INVESTIGATION OF THE STRUCTURE OF NN'-DIGLYCYL-L-CYSTINE

A) The determination of the trial structure

1. Interpretation of the hOL Patterson projection

The unsharpened hOL Patterson for NN'-diglycyl-L-cystine was computed from Beavers-Lipson strips. It is shown in Figure III on an arbitrary scale.

It has been mentioned that the space group requires the diglycyl cystine molecules to lie in special positions and it was suggested that, as the peptide itself might be expected to show a two-fold axis of symmetry, a likely arrangement would place the molecules across the two-fold axes of the cell. By referring to the sketch on page 25, it can be seen that such a two-fold axis must intersect the S-S bond which in turn would have to be parallel to the hOL plane of projection. The Patterson interaction peak arising from the S-S distance would then be found at some point on a circle of radius equal to the S-S bond length described about the origin of the hOL projection.

With 2.08A. as the radius, this being twice the value given by Pauling (29) for the single covalent bond radius of sulfur, the circle was drawn and the indicated peak selected as the S-S interaction. There was little ambiguity in this selection as no other areas on or near this circle showed the presence of a large peak such as would be expected for two sulfur atoms.

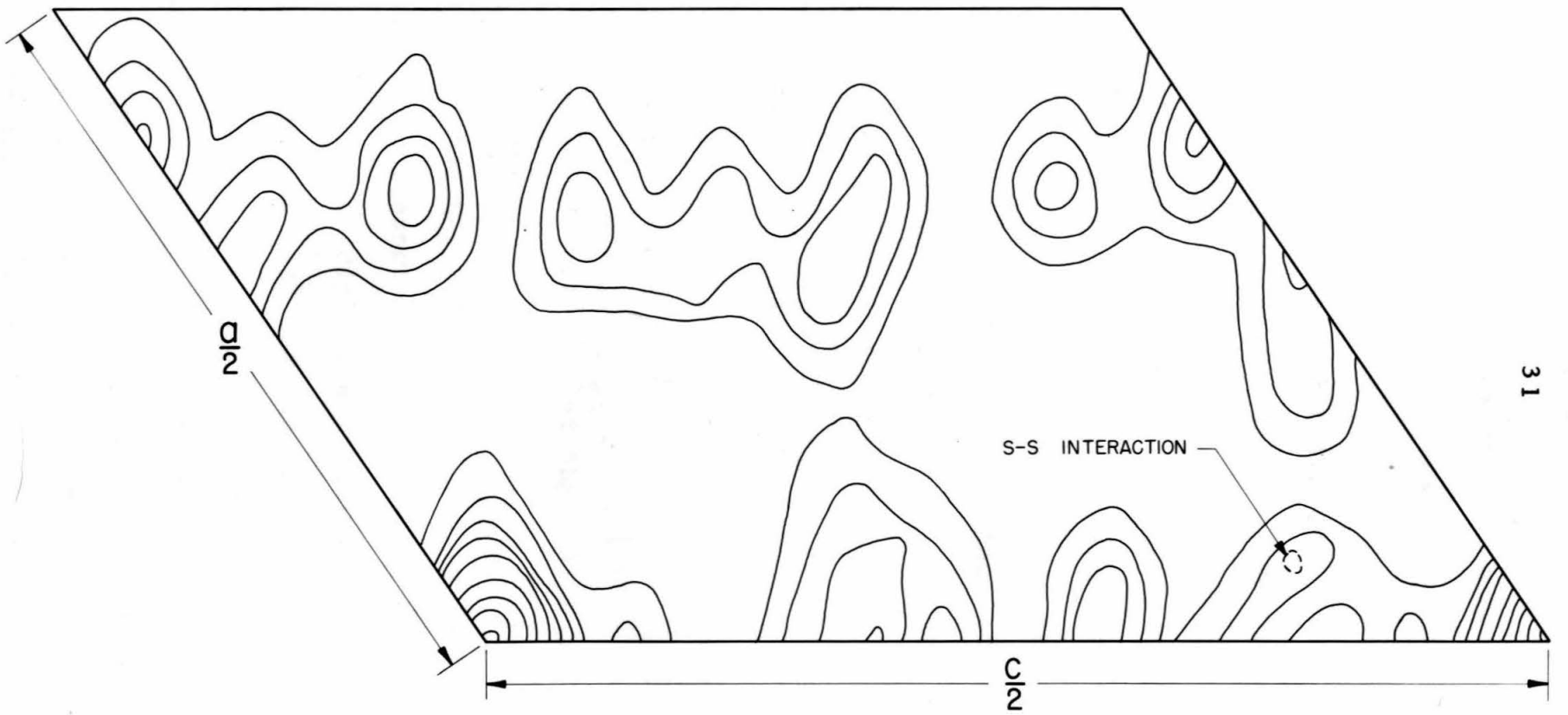


Fig.III. hOl PATTERSON PROJECTION NN'-DIGLYCYL-L-CYSTINE

One of the other noteworthy features of this Patterson was the row of very strong peaks along the line u = 1/3. These suggested that a chain of atoms might lie approximately along the line x = 1/3 in the unit cell. The observed peaks would then be produced by interactions involving a sulfur atom and the atoms constituting the chain. (This concept is also in agreement with the unusually high intensity of the 300 reflection.) The picture of a row of atoms more or less parallel to c in projection received further support from a comparison of cell dimensions for this structure with those found for β-diglycine by Hughes (8). These two sets of data are given in the table below.

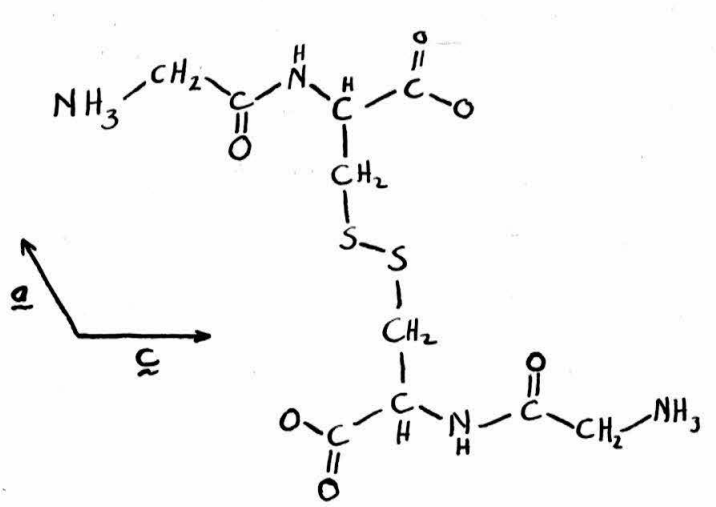
Compound	a(kX)	b(kX)	c(kX)	β	Z	Space Group
NN'-Diglycyl-L-cystine	12.24	4.83	17.14	124°24'	2	A2
β-Diglycine	17.89	4.62	17.06	125°10'	8	A2/a

It must be kept in mind that although there are only 2 diglycylcystine molecules per cell, these lie across twofold axes and thus the Patterson should be interpreted as indicating the presence of two chains per molecule or four per cell. In the β-diglycine structure there are eight diglycine chains per cell and, as was shown by Hughes, these are oriented parallel to c. Then to allow for the difference in space group, the a axis of β-diglycine should be halved before comparing the dimensions. If this is done, several striking similarities are noticeable. Firstly, the two c axes are almost identical, as are the two β angles, indicat-

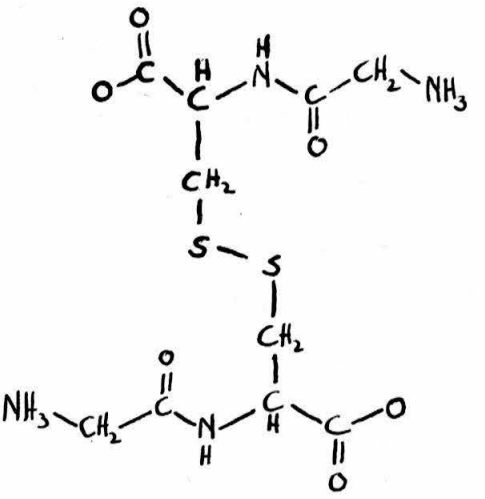
ing that the chains present in NN'-diglycyl-L-cystine are probably identical with ~~the~~ diglycine chains and are probably oriented in the same direction as those in the β-diglycine structure. Secondly, the longer b axis reported for diglycylcystine can be interpreted as due to a greater tipping of the chain out of the a - c plane than was found in β-diglycine. (It will be shown later that van der Waals repulsion between the atomic configurations connecting the chains also accounts for part of this increase in b.) Thirdly, the a axes are the only ones which are radically different, that for diglycylcystine being 3.29 kX larger than the halved a axis of β-diglycine. This was ascribed to the presence of atoms interconnecting the chains in diglycylcystine, only van der Waals forces separating the chains in β-diglycine.

The general view of the structure, as derived from the Patterson projection and this comparison of cell dimensions, was, therefore, of diglycine chains roughly parallel to c interconnected by means of some kind of bridge containing the sulfur atoms of the molecule, presumably a "di-sulfide bridge".

Again referring to the sketch on page 25, two likely trial structures obeying all the aforementioned criteria can be proposed at once. In projection, they are:



I.



II.

In both models it is seen that the S-S bond is in the correct position as shown by the Patterson and there are diglycine chains oriented parallel to c and at roughly x = 1/3.

Synthetic Patterson functions for only intermolecular interactions were constructed for each model. This was done simply by drawing a scale projection of the particular model and then, with each atom in turn as the origin and with the orientation of the molecules constant, noting the positions of the ends of vectors from that atom to others in the same molecule. On the basis of these synthetic Pattersons it was not possible to choose between models I and II since both fit the calculated Patterson projection equally well.

The elimination of one or the other of the above models was therefore accomplished by actual calculation of structure factors. x and z parameters of the atoms were estimated from scale projections and the atom form factors for each hOL

plane obtained from Hartree curves. Hydrogens were not included in the calculation at this stage. The temperature factor $e^{-B\frac{\sin^2\theta}{\lambda^2}}$ with B = 3.5, predicted for the hOL zone by Wilson's method, was applied to the individual form factors. As the hOL zone of the A2 space group is centro-symmetric, structure amplitudes instead of intensities were calculated and compared with the absolute square roots of the corrected observed intensities. Bragg-Lipson charts were used for the actual computation. The water molecules were placed on the two-fold screw axes in this work. Comparison of calculated structure factors with observed data indicated fairly good agreement for low order reflections for model II, but only random agreement for model I. It was therefore assumed that model II was essentially correct in projection, and work on Fourier refinement of parameters was begun.

Before leaving this phase of the investigation, it would be well to state that still another criterion was used to check the acceptability of model II as against model I. It was noted that both trial models had all atoms far enough away from the line x = 1/2 to insure that the electron density along that line would be approximately zero. Then:

$$\rho(1/2,z) = \frac{1}{A}\sum_h\sum_\ell F_{h0\ell}\cos 2\pi(h/2+Lz) = 0 \qquad 1.$$

$$\sum_h\sum_\ell F_{h0\ell}\cos\pi h\cos 2\pi Lz = 0 \qquad 2.$$

Multiplying by $\cos 2\pi L'z$, where L' is some particular value of L, and using the Fourier integral theorem,

$$\sum_h\sum_\ell F_{h0\ell}(-1)^h\int_0\cos 2\pi Lz\cos 2\pi L'z\,dz = 0 \qquad 3.$$

For L = L' this integral has the value 1/2 so that:

$$\sum_h F_{hol}(-1)^h = 0 \text{ for } L = L' \qquad 4.$$

It was thus possible to substitute the square roots of a series of observed intensities into this relationship and to give those which agreed in magnitude with the calculated structure factors the signs of the latter quantities. One could then get an idea of how the signs of the remaining amplitudes would have to be distributed in order to satisfy equation 4. For model II it was noted that in a majority of cases the signs thus selected agreed with the signs of the calculated structure factors even though the magnitudes were not in agreement. For model I, however, many cases arose in which the signs of the calculated structure factors were opposite to the signs predicted by use of equation 4. This again lent support to model II over model I.

It should be mentioned that application of equation 4 was also useful in selecting terms for the preliminary Fourier syntheses. Thus it often happens that in these Fouriers terms were included whose calculated structure factors, although not in good agreement with the observed amplitudes, did have signs which tended to satisfy equation 4.

2. The preliminary refinement of the x and z parameters

Due to the relatively short b axis, it was seen that the estimation of y parameters for this structure would be a rather difficult task. As is often the case in such problems, it was decided to work first with the hOL zone reflections

exclusively and to determine the y parameters only after the x and z parameters had been refined as far as the projection data would allow. This section will describe the refinement using i) Fourier syntheses, and ii) two-dimensional least squares.

The structure factor calculation used to decide between possible trial structures I and II was now used to determine which reflections in the hOL zone might profitably be included in a Fourier projection. 57 spectra for which the agreement between observed and calculated structure factors was good were employed in this first Fourier. The inclusion of even 57 terms was perhaps over-optimistic at this juncture because the Fourier peaks were but poorly resolved. One feature which did appear in this projection, however, was the shift of the peak representing the oxygen of the water molecule from x = 0, z = 1/4 to a position off the two-fold screw axis. This of course would mean that the form being examined was actually a dihydrate instead of a monohydrate as reported by Greenstein. It has been mentioned that moving the water molecules off these axes would be to improve the O -H --O hydrogen bond distances and the H -O --H angles. Therefore this result obtained from the first Fourier was accepted without much doubt even though the rest of the projection was of little use in estimating atomic parameters.

Placing the water of hydration at the position roughly indicated by this preliminary Fourier, a new structure factor calculation for hOL reflections was made. International

Business Machines were used for these computations. (See Appendix II for a description of the methods used.) A comparison with observed structure factors was again made and, using a more strict agreement as a basis for selection, 45 reflections chosen for inclusion in a second Fourier projection. Wherever possible equation 4 was used to check the proposed signs of the observed structure factors. This Fourier is shown in Figure IV. The projection clearly demonstrates that the decision to move the water molecule away from the screw axis was indeed correct. Moreover, it is also apparent that the nitrogen atom of the amine group has been incorrectly placed and must be raised. It should be noted that the carboxyl carbon was not resolved from one of the oxygen atoms to which it is linked, as was also the case with the carbon and oxygen atoms of the amide link. Certain unresolved atoms were to be expected, however, since the 4.8A. repeat in the b direction indicates that the peptide chains are tipped at a considerable angle to the a-c plane.

The atoms were now placed as close to the maxima of the Fourier peaks as possible and new parameters recorded. No back shift rules such as had been used in the L-threonine structure were applied. In the case of unresolved atoms the new parameters were estimated as carefully as possible from the shape and magnitude of the Fourier peaks.

A structure factor calculation with the revised parameters was made and, on comparing observed with calculated data, it was seen that 75 reflections could now be safely

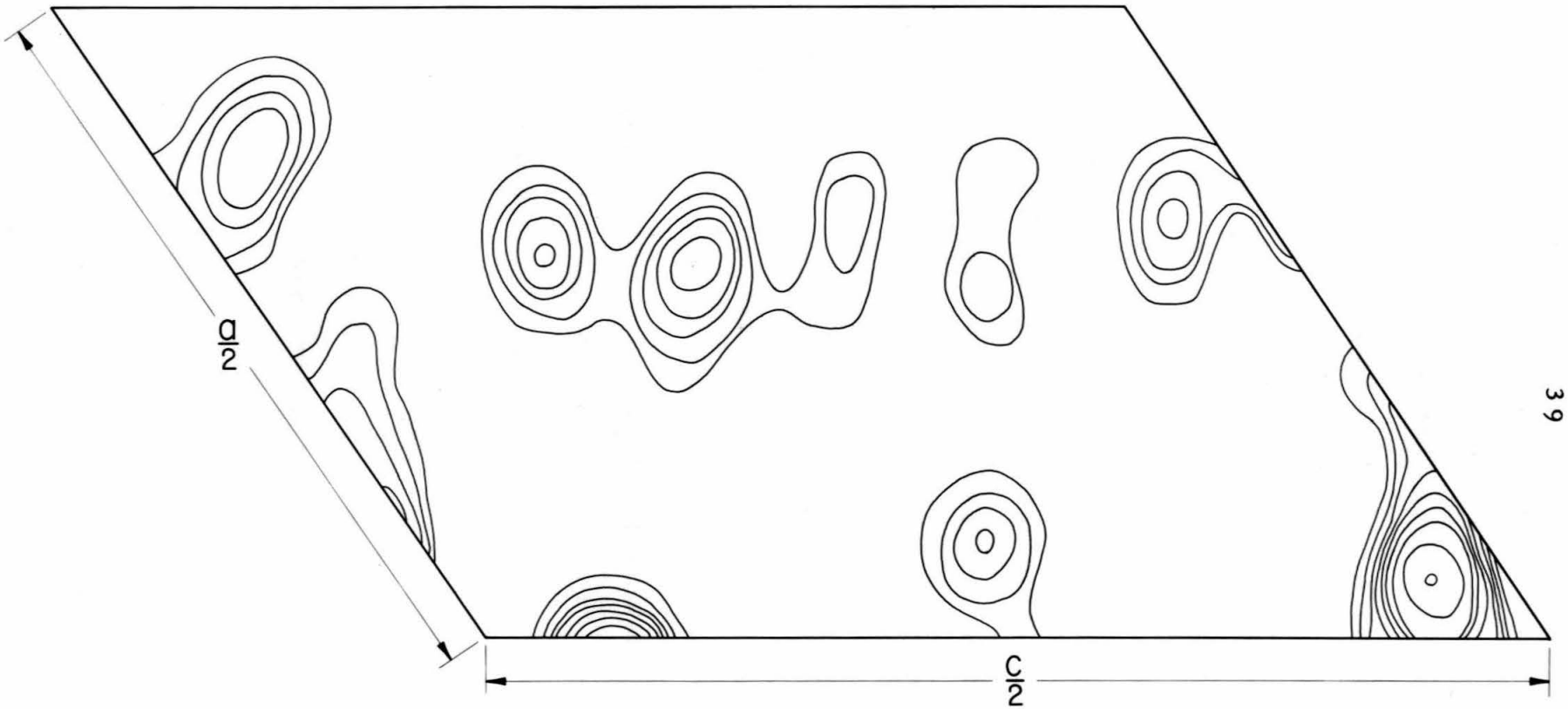


Fig. IV. PRELIMINARY hOl FOURIER USING 45 TERMS NN'-DIGLYCYL-L-CYSTINE

included in a third Fourier synthesis. The projection, when computed, again showed the distorted amine nitrogen peak remarked in the second Fourier. A better resolution of the atoms was also noted.

Again the peak maxima were chosen as the centers of the respective atoms and parameters estimated. These parameters are given in the first column of Table I. Note that two sets of parameters for the amine nitrogen are listed corresponding to the ambiguity of the Fourier peak for that atom. Two separate structure factor calculations, one for each set of amine nitrogen parameters, were made and the results compared with observed data. It was obvious that structure factors computed with parameter set 2 for the nitrogen in question were in closer agreement with observation than were those with parameter set 1.

A fourth Fourier was now prepared with 134 hOL reflections for which agreement was sufficiently good. One exception to this statement should be noted, however. The 300 reflection, by far the most intense one obtained from the crystal, was assigned an observed amplitude of 110 on an absolute scale while the calculated structure factor for this plane was consistantly around 170. As there was doubtless a secondary extinction effect reducing the observed intensity, the reflection was put into this new Fourier with the amplitude 170. No direct calculation of the amount of secondary extinction to be expected for this spectrum was made, but the general good agreement of the other hOL zone reflections

made it unlikely that this one reflection was so badly out of agreement due to parameter errors alone and justified the use of the calculated structure factor in place of the observed. Complete resolution and a single unambiguous peak for the amine nitrogen was obtained in this fourth projection. The overlapping in the carboxyl and carbonyl groups was still present, however. Parameters obtained from this Fourier are given in the second column of Table I. It is seen that the parameters from the fourth Fourier differ by about 0.1A. from the third Fourier parameters. These shifts are rather large, but then less than half the observed intensities were included in Fourier III, while over 80% were used in the calculation of Fourier IV.

A complete reestimation of all hOL zone intensities was now made from a set of Weissenbergs of longer exposure than had previously been used. This was done in an effort to observe more of the weak spectra in the range where the Lorentz and polarization factors are near unity. Results of the measurement were compared with the old data and it was seen that when the two sets were placed on the same scale differences no greater than 10% occurred for reflections measurable in both. Because of this rather good agreement the same absolute scale and temperature factors were used in treating the new data.

Structure factors computed with parameters from Fourier IV showed very good overall agreement with the square roots of these new intensities. It was now possible to calculate

a Fourier projection using all 160 non-zero observed intensities in the hOL zone. This Fourier is shown in Figure V while the parameters obtained therefrom are given in the third column of Table I. Note that the parameter differences between Fourier IV and V correspond to shifts of only about 0.05A. as against 0.10A. for the corresponding difference between Fouriers III and IV. Due to the lack of resolution, the parameters of C_1, O_2, C_3 and O_3 were undoubtedly subject to larger errors than those of the other atoms.

With these new parameters, still another structure factor calculation for the hOL zone was made. However, the distribution of signs did not change at all from that used to compute Fourier V. It was therefore assumed that the limit of Fourier refinement in this projection had been attained and that recourse to least squares refinement would have to be made. At this stage in the investigation the "reliability factor", defined by the equation

$$R = \frac{\Sigma |F_{obs} - F_{calc}|}{\Sigma |F_{obs}|} \cdot 100$$

was 22.4% for the hOL zone.

The general mathematical background of the least squares method can be found in a number of standard works, notably in "The Calculus of Observations" by Whitaker and Robinson (30). The manner of application of this method to the problems of crystal structure analyses has been presented by Hughes (31), Shoemaker, Donohue, Schomaker, and Corey (3), etc. A discussion of the errors in the results of least squares solutions

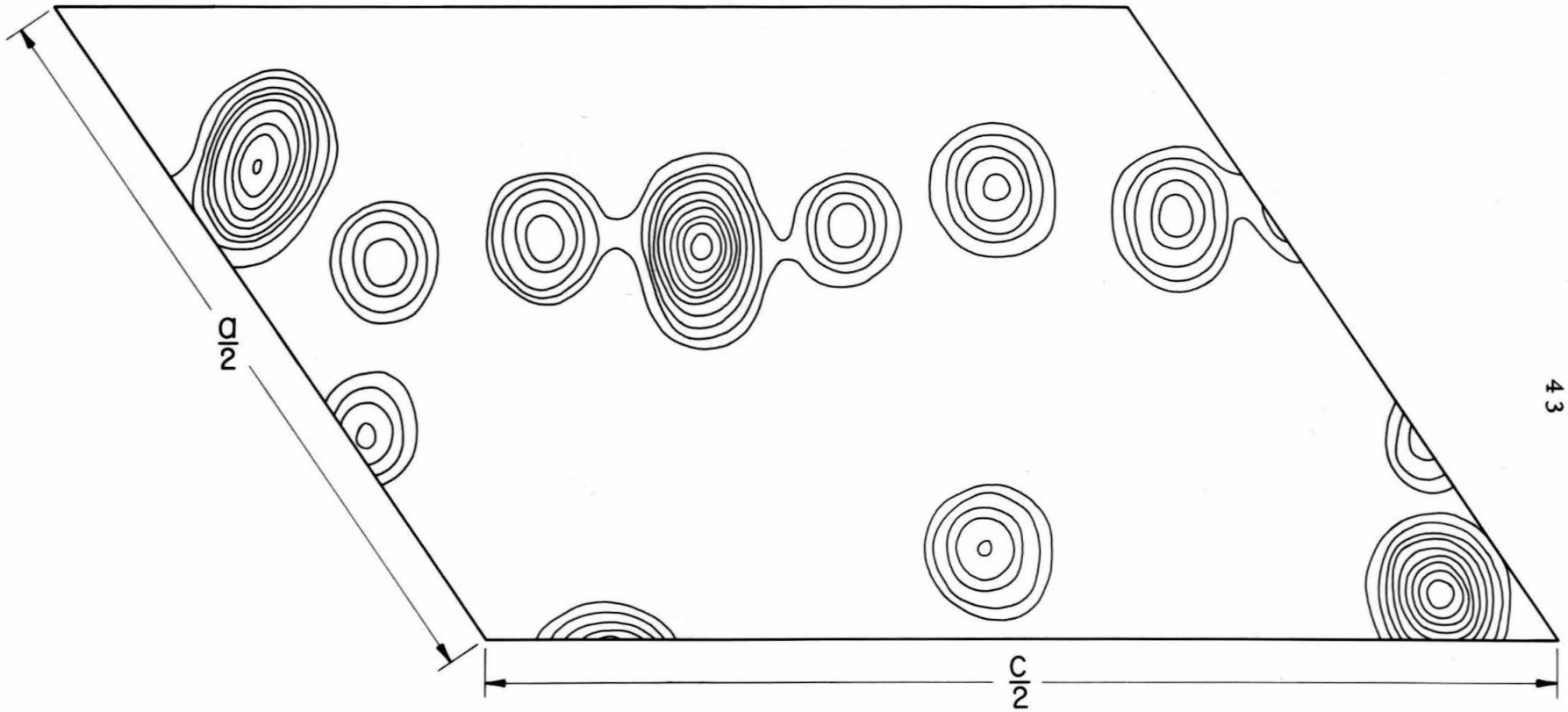


Fig. V. FINAL h0l FOURIER USING 160 TERMS NN'-DIGLYCYL-L-CYSTINE

has been given by Cruikshank (32).

Briefly, if one desires to find the most probable answers for x and y from a set of equations:

TABLE I

PARAMETERS FROM hOL FOURIERS

	Fourier III.		Fourier IV.		Fourier V.	
Atom	x	z	x	z	x	z
C_1	0.342	0.022	0.340	0.025	0.344	0.026
C_2	.278	.060	.290	.068	.294	.073
C_3	.320	.230	.320	.223	.320	.228
C_4	.327	.305	.327	.300	.325	.298
C_5	.167	.005	.163	.006	.160	.007
N_1	.310	.158	.315	.152	.316	.155
(1)N_2	.275	.348				
(2)N_2	.343	.375	.345	.377	.351	.378
O_1	.337	.455	.332	.455	.330	.455
O_2	.395	.052	.392	.052	.389	.055
O_3	.280	.200	.298	.214	.300	.217
O_4	.079	.260	.079	.262	.075	.260
S	.037	.457	.035	.455	.036	.457

$$u_1 = f_1(x,y), \quad u_2 = f_2(x,y), \quad ----, \quad u_s = f_s(x,y) \qquad 1.$$

where f_1----f_s are known functions and u_1----u_s are experimentally measurable quantities subject to accidental error, one can start with an approximate pair of values, $\bar{x}$ and $\bar{y}$, and put:

$$x = \bar{x} + \Delta x, \quad y = \bar{y} + \Delta y \qquad 2.$$

Then:

$$f_1(x,y) = f_1(\bar{x}+\Delta x, \bar{y}+\Delta y)$$
$$= f_1(\bar{x}, \bar{y}) + \frac{\partial f_1(\bar{x}, \bar{y})}{\partial \bar{x}} \Delta x + \frac{\partial f_1(\bar{x}, \bar{y})}{\partial \bar{y}} \Delta y \quad 3.$$

It should be emphasized that if Δx and Δy are not true infinitesimals this last equation is only approximately true. In such a situation, as stated by Hughes (31), one least squares solution will not be sufficient to obtain accurate values of Δx and Δy, but the process must be repeated several times.

Written as:

$$\begin{aligned} \frac{\partial f_1(\bar{x}, \bar{y})}{\partial \bar{x}} \Delta x + \frac{\partial f_1(\bar{x}, \bar{y})}{\partial \bar{y}} \Delta y &= u_1 - f_1(\bar{x}, \bar{y}) \\ \cdots\cdots\cdots\

minimization of the sums of the squares of the errors requires that the following equations be satisfied:

$$[aa]\Delta x + [ab]\Delta y = [an]$$
$$[ab]\Delta x + [bb]\Delta y = [bn] \qquad 6.$$

where $n_1 = u_1 - f_1(\bar{x}, \bar{y})$ and $[an] = a_1 n_1 + a_2 n_2 + \cdots$.

Equations 6 are called the normal equations and it is easily seen that there are as many as there are variables in the observational equations. Therefore a unique solution for the variables can be obtained. It should also be noted that the various observational equations can be given different weights in the least squares method corresponding to different estimated degrees of reliability of experimental results.

As applied to crystal structure work, the quantities u_1 --- u_s are usually either observed absolute intensities or their square roots with proper signs. The functions f_1 ---- f_s are the structure factor functions for each reflection from the entire crystal or from a particular zone if two-dimensional least squares are being calculated. The parameters $\bar{x}$, $\bar{y}$, etc., are the parameters from a trial structure or from a preliminary Fourier refinement, while the variables Δx, Δy, etc., are the unknown coordinate shifts which will tend to minimize the differences between calculated and observed structure factors or intensities. Of course Δx, Δy, etc. are not true infinitesimals so that, as has been stated above, one least squares calculation will not give the best possible parameter shifts.

For the A2 space group the equations for the structure factors are:

$$F^2 = A^2 + B^2 \qquad 7a.$$

$$A = \sum_i 4f_i \cos 2\pi(hx_i + Lz_i) \cos 2\pi k y_i \qquad 7b.$$

$$B = \sum_i 4f_i \cos 2\pi(hx_i + Lz_i) \sin 2\pi k y_i \qquad 7c.$$

In the centrosymmetric hOL zone amplitudes instead of intensities may be used so that the observational equations are of the form:

$$\sum_i \sqrt{w_{hOL}} \left(\partial F_{hOL}/\partial q_i \cdot \Delta q_i\right) = \sqrt{w_{hOL}} \cdot \Delta F_{hOL} \qquad 8.$$

where $\sqrt{w_{hOL}}$ is a weighting factor for the reflection in question, q_i is a generalized coordinate of the i^{th} atom and $\Delta F_{hOL} = F_{obs} - F_{calc}$. Only non-zero F_{obs} are usually incorporated in these calculations.

Restricting the calculation to the hOL zone only derivatives with respect to x and z can be computed. The parameter shifts derived from such a two-dimensional least squares must not be considered as final since a great mass of data from the h1L, h2L, and h3L zones is neglected.

The respective derivatives for this particular problem have the form:

$$\frac{\partial F_{hOL}}{\partial x_i} = -8\pi h f_i \sin 2\pi(hx_i + Lz_i)$$

$$\frac{\partial F_{hOL}}{\partial z_i} = -8\pi L f_i \sin 2\pi(hx_i + Lz_i) \qquad 9.$$

The actual calculations for the first two-dimensional least squares on diglycylcystine were performed at the same

time that the structure factor calculation using parameters from Fourier V was made. (See Appendix II). The weighting factors were obtained from the following equations:

$$w_{hOL} = 100/F^2_{hOL} \quad \text{for } F^2_{hOL} \geqslant 16\ F^2_{hOLmin.} = 100$$

$$w_{hOL} = 1 \quad \text{for } F^2_{hOL} \leqslant 16\ F^2_{hOLmin.} \qquad 10.$$

This is essentially the same weighting scheme as used by Hughes (31) in the investigation of the structure of melamine and is equivalent to taking the percentage error in the intensities as constant for intensities more than sixteen times the background and the absolute error constant for lesser intensities.

After calculating the coefficients of the observational equations and the Δ F terms, the coefficients of the normal equations were computed. (See Appendix II.) Here again the observed intensity of the 300 reflection was discounted and the calculated value inserted in its place. There being one half a molecule of formula $C_{10}N_4O_6S_2H_{18}\cdot 2H_2O$ per asymmetric unit of the structure, or twelve atoms excluding hydrogens, twenty four normal equations were derived. They were solved in a reasonably short period by the iteration method of Gauss and Seidel described in "The Calculus of Observations". The resultant coordinate shifts are listed in column two of Table II while the corrected parameters are given in the third column.

Comparing the shifts obtained from the results of Fouriers IV and V with those from the first least squares,

it is seen that several large changes occur in the latter. This is particularly true of the parameters of C_1, C_3 and O_3. However these are just the atoms left unresolved by the hOL Fourier projection so it might be expected that their parameters would change markedly when the greater resolving power of the least squares method was applied to the problem.

Parameters from the first least squares solution were now used to recompute the hOL structure factors. Tabulation of the results, after recalculation of the absolute scale and temperature factors, showed that the reliability factor had decreased to 17.4%. (It is interesting to note that even though the parameter shifts themselves were large, and therefore the probable errors in the corrected coordinates were large, the discrepencies between observed and calculated amplitudes were much smaller. To the author's knowledge this property of least squares solutions as employed in structure work has not been previously reported.)

Again, while calculating structure factors the coefficients of observational equations were calculated preparatory to the formation of a second two-dimensional least squares. The derivation of the normal equations was accomplished as before: solution of them gave the parameter shifts listed in the fourth column of Table II and the amended parameters in column five.

The coordinate changes obtained from the second least squares can be seen to be smaller on the average than the first least squares shifts. The probable errors in the re-

fined parameters were not calculated specifically but from the magnitude of the shifts they could be estimated as about 0.02 to 0.03A. A structure factor calculation using the new parameters showed that the agreement between observed and calculated data was now about 16.6%. This does not represent a large decrease in the reliability factor, but the corrected parameters undoubtedly have much smaller probable errors than the previous parameter set.

Since there was little point in continuing the refinement of the x and z parameters beyond this stage without using complete three dimensional data, an attempt was now made to estimate the y parameters. This phase of the investigation and the subsequent refinement of parameters will be presented in the following section.

TABLE II.

	Parameters from Fourier V		Shifts from 1st Least Squares		Corrected Parameters	
Atom	x	z	x	z	x	z
C_1	0.344	0.026	-0.004	-0.005	0.340	0.021
C_2	.294	.073	.000	-.004	.294	.069
C_3	.320	.228	.010	.004	.330	.232
C_4	.325	.298	-.003	-.001	.322	.297
C_5	.160	.007	.003	.000	.163	.007
N_1	.316	.155	-.002	-.001	.314	.154
N_2	.351	.378	-.001	-.002	.350	.376

(cont'd. next page)

TABLE II (CONT'D.)

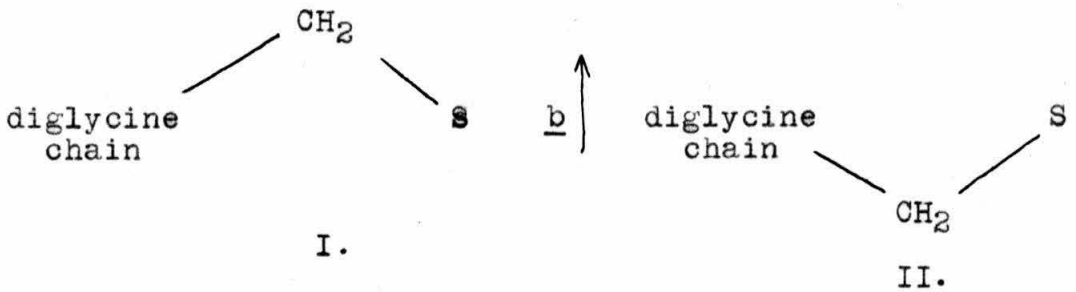
	Parameters from Fourier V		Shifts from 1st Least Squares		Corrected Parameters	
Atom	x	z	x	z	x	z
O_1	.330	.455	.000	-.001	.330	.454
O_2	.389	.055	.000	-.001	.389	.054
O_3	.300	.217	-.007	-.002	.293	.215
O_4	.075	.260	-.003	.001	.072	.261
S	.036	.457	-.004	-.001	.032	.456

	Shifts from 2nd Least Squares		Corrected Parameters	
Atom	x	z	x	z
C_1	0.001	0.002	0.341	0.023
C_2	-.004	-.002	.290	.067
C_3	-.006	-.003	.324	.229
C_4	.002	.000	.324	.297
C_5	-.003	-.001	.160	.006
N_1	.004	.001	.318	.155
N_2	.000	.000	.350	.376
O_1	.001	.000	.331	.454
O_2	.001	.002	.390	.056
O_3	.001	.000	.294	.215
O_4	-.004	-.002	.068	.259
S	.000	.000	.032	.456

B) The completion of the structure analysis and final refinement of parameters

1. An estimation of y parameters aided by a calculation of several line Pattersons and Patterson sections.

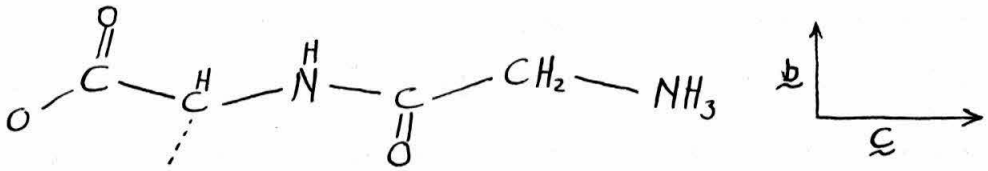
By starting with the refined x and z parameters obtained from projection work, the y parameters were estimated from i) the distances between bonded atoms in projection, and ii) the usually accepted bond lengths for peptide molecules (11). It was seen at once that certain ambiguities would arise in such a method. For example, assuming the y parameter of the sulfur atom in the asymmetric unit to be zero, this being allowed in the non-centrosymmetric A2 space group, the bond from that sulfur to the adjacent methylene carbon atom might be oriented in either of the following ways:



I. II.

From the hOL projection alone it was almost impossible to choose between these two configurations. Then again, the orientation of the diglycine chain with respect to the b axis might at first glance be thought uncertain. On closer inspection, however, this was not found to be the case. Considerations of the necessary coplanarity of the amide group

and hydrogen bonding conditions supported the following model for the chain:



It would appear that this model of the diglycine chain would settle the doubt concerning the $-CH_2-S-$ side chain configuration in favor of model II. Actually, this was by no means obvious when ball and stick models of the molecule were prepared and superimposed on the hOL projection, certain rotations about the C - C and C - N bonds giving a plausible fit for the model constructed using configuration I.

Still another ambiguity was noted when the y parameters of the water oxygen were estimated. The water molecule must be forming two hydrogen bonds, one to the water molecule immediately beneath or above it and related to it by the two-fold screw rotation, and the other to one of the oxygen atoms of the diglycine chain. It is this latter bond which would fix the roughly estimated y parameter of the water oxygen. Examining the projection, it was noted that this oxygen, by suitable changes in its y parameter, could be brought close enough to either a carboxyl oxygen or the amide carbonyl oxygen to permit hydrogen bond formation. Since the projection distances were, in both cases, about what would be expected for an O-H---O bond the y parameter of the water oxygen could be taken equal to the y parameter of either chain oxygen as a

first approximation.

To resolve these ambiguities Patterson sections and line Pattersons were employed. Three dimensional intensity data were obtained from the h1L, h2L, h3L, hk0 and hk1 Weissenbergs. This was probably not far from complete data, only a few weak h4L reflections being unobserved. The corrected hk0 intensities were put on an approximate absolute scale by comparison of the observed h00 spectra with the absolute h00 intensities from the h0L zone. Similarly, an absolute scale factor for the h2L zone was obtained by comparison with absolute h20 intensities. Scale factors for the h1L and h3L zones were computed by Wilson's method (27).

Two independent sets of y parameters for the peptide molecule were now estimated from the projection distances and usual bond lengths, the first set from model I of the $-CH_2S-$ side chain and the second from model II. The y parameters of the chain atoms differed widely from set to set so that the line Pattersons might be expected to give a clear-cut solution. The two parameter sets are given in Table III.

These parameters are for the peptide molecule with y_S arbitrarily chosen as 0.000: no water oxygen is given since the uncertainty as to its position could not be clarified until the uncertainty as to the side chain configuration had been resolved.

TABLE III

	Model I.	Model II.		
Atom	y	y'	x	z
C_1	0.300	0.072	0.341	0.023
C_2	.109	.891	.290	.067
C_3	.145	.908	.324	.229
C_4	.353	.116	.324	.297
C_5	.277	.723	.160	.006
N_1	.237	.019	.318	.155
N_2	.171	.934	.350	.376
O_1	.175	.937	.331	-.046
O_2	.524	.286	.390	.056
O_3	.893	.653	.294	.215
S	.000	.000	.032	-.044

The z parameters of the sulfur atom, O_1, O_2, C_1, C_2, and C_5 are so nearly equal that a Patterson section at w = 0 was calculated in the expectation that several, if not all, these vectors would be resolved. This section is shown in Figure VI. The appearance of peaks in positions predicted by the y parameters of model II rather than model I was interpreted as strongly indicative of the correctness of the former.

To confirm this conclusion line Pattersons at values of the vector coordinates u and w corresponding to other sulfur - chain atom distances were computed. These also served to support the configuration depicted by model II.

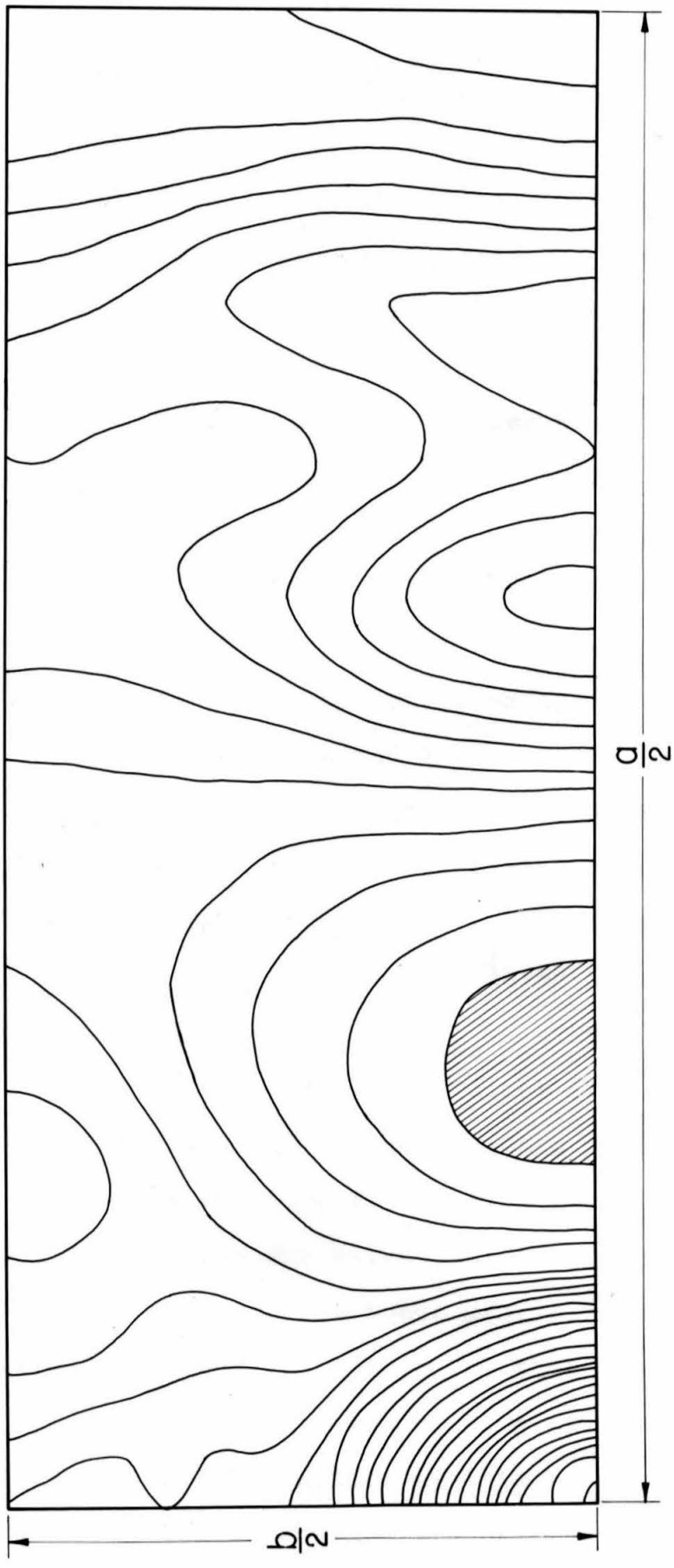


Fig. VII. PATTERSON SECTION AT w=0 FOR NN'-DIGLYCYL-L-CYSTINE

Having established the side chain configuration, an attempt was now made to fix the y parameter of the water oxygen by means of line Pattersons at u = 0.036, w = 0.197 and at u = 0.100, w = 0.215 along which one would expect to find peaks representing the interactions of this atom with the sulfur atom at y = 0.000. These Pattersons, when computed, showed clearly that the water oxygen was at nearly the same level as the amide carbonyl oxygen and was therefore forming a hydrogen bond to that atom rather than to O_1. This result was anticipated necessarily beforehand since if the hydrogen bond had been directed to O_1 this oxygen would have been accepting three such bonds, a situation never before encountered.

Structure factors A and B calculated for the h1L, h2L, and h3L zones from the rough parameters of set 2 showed fair agreement with observed intensities. From these structure factors a recalculation of scale and temperature factors in the three zones was made using a graphical averaging method which plots $\Sigma F^2_{calc}/\Sigma F^2_{obs}$, summed over a given range of sin θ, vs: sinθ. A straight line drawn through the scatter of points then gives the scale factor from its intercept and the temperature factor from its slope. The results of this calculation suggested that the temperature factors for the h1L, h2L, and h3L zones were 3.8, 4.2 and 4.7A^2 respectively. No obvious reason for this trend could be found so a uniform temperature factor of 4.0A^2 was applied to the calculated structure factors at this stage.

A one dimensional y least squares was now prepared as

the first step in the refinement of the y parameters. Due to the complex structure factor, this least squares solution had to be performed on the F^2 values rather than on F's.

$$F^2 = A^2 + B^2$$

$$\frac{\partial F^2_{hkl}}{\partial y_i} = 2A \frac{\partial A}{\partial y_i} + 2B \frac{\partial B}{\partial y_i} \qquad 11.$$

From the definition of A and B given in equation 7:

$$\frac{\partial A}{\partial y_i} = -4.2\pi k f_i \cos 2\pi(hx_i + Lz_i) \sin 2\pi k y_i$$

$$= -2\pi k B_i$$

$$\frac{\partial B}{\partial y_i} = 4.2\pi k f_i \cos 2\pi(hx_i + Lz_i) \cos 2\pi k y_i$$

$$= 2\pi k A_i$$

Therefore:

$$\frac{\partial F^2_{hkl}}{\partial y_i} = 4\pi k(BA_i - AB_i) \qquad 12.$$

By neglecting the x and z parameters in this least squares one is in effect neglecting the Δx and Δz shifts in comparison with the Δy shifts. This was probably a good approximation for at least the preliminary y refinement.

The weights of the observational equations were defined as:

$$\sqrt{w_{hkl}} = 100/\ F^2_{hkl} \quad \text{for } F^2_{hkl} \geq 16F^2_{hkl\,min.} = 100$$

$$\sqrt{w_{hkl}} = 1 \quad \text{for } F^2_{hkl} \leq 16F^2_{hkl\,min.} \qquad 13.$$

The observational equations were therefore of the form:

$$\sum_i \sqrt{w_{hkl}} \frac{\partial F^2_{hkl}}{\partial y_i} \Delta y_i = \sqrt{w_{hkl}} \cdot \Delta F^2_{hkl} \qquad 14.$$

The derivatives and ΔF^2 terms were computed from the structure factor calculation mentioned above. Only h1L, h2L and h3L data were used since the derivatives with respect to y for h0L reflections were, of course, zero.

Normal equation coefficients were calculated in the usual manner and the twelve equations solved for the parameter shifts Δy_1 -----Δy_{12}. The changes and the corrected parameters appear in the second and third columns of Table IV.

Another structure factor calculation for the h1L, h2L and h3L zones was now made using the improved y coordinates. The agreement between observed and calculated intensities was significantly better than it had been for the prior calculation. Again the scale and temperature factors for each zone were computed by the graphical averaging method already discussed. The results indicated that the scale factors for all zones were essentially correct but that the temperature factors were in error. The revised temperature factors for h1L, h2L and h3L reflections were 3.8, 4.2 and 4.7A^2 in complete agreement with the results obtained from the first such calculation. Thus it would seem that there actually is a gradual increase in B with increasing k. This cannot be laid to an anisotropic temperature factor but presumably must be due to some as yet unexplained absorption or extinction effect. Even though no theoretical reason for the result could be advanced, these empirical values of the temperature factor were applied to the calculated structure factors.

Observational and normal equation coefficients for a second y least squares were derived from the structure amplitudes A and B after the emendation noted above. The resultant parameter changes and the corrected y parameters are presented in Table IV.

Having with this step concluded the preliminary refinements of the x, y, and z parameters, three dimensional least squares were now employed to complete the analysis. A description of these calculations and their results, together with a discussion of the errors in the final parameters, will be given in the next few pages.

TABLE IV.

Atom	Preliminary y parameters	Shifts from 1st Least Squares	Corrected Parameters
C_1	-.072	-0.002	0.042
C_2	.891	-.002	.871
C_3	.908	-.010	.880
C_4	.116	-.006	.092
C_5	.723	.008	.713
N_1	.019	-.002	.999
N_2	.934	-.006	.910
O_1	.437	.009	.428
O_2	.286	-.003	.265
O_3	.653	.009	.642
O_4	.646	.000	.628
S	.500	.018	.500

TABLE IV (CONT'D.)

Atom	2nd Shifts from Least Squares	Corrected Parameters
C_1	0.000	0.024
C_2	-.007	.846
C_3	-.011	.851
C_4	-.008	.066
C_5	.015	.710
N_1	.006	.987
N_2	-.002	.891
O_1	.011	.421
O_2	-.001	.246
O_3	-.003	.621
O_4	-.010	.600
S	.018	.500

2. The refinement of the structure using three-dimensional least squares methods.

A general three dimensional least squares on a structure as complicated as NN'-diglycyl-L-cystine has not been reported in the literature. The total of thirty six unknown parameter shifts to be obtained from only about 800 observed intensities suggested that the resultant parameters might be subject to fairly considerable errors necessitating the calculation of several such least squares solutions in order to obtain satisfactory refinement.

The observational equation coefficients for the general hkl reflections were derived as follows:

$$\sum_i \sqrt{w_{hkl}} \frac{\partial F^2_{hkl}}{\partial q_i} \cdot \Delta q_i = \sqrt{w_{hkl}} \cdot \Delta F^2_{hkl} \qquad 15.$$

$$\frac{\partial F^2_{hkl}}{\partial q_i} = 2A \frac{\partial A}{\partial q_i} + 2B \frac{\partial B}{\partial q_i} \qquad 16.$$

Using equations 7a and 7b:

$$\begin{aligned}
\frac{\partial A}{\partial x_i} &= -8\pi h f_i \sin 2\pi(hx_i + Lz_i) \cos 2\pi k y_i \\
\frac{\partial B}{\partial x_i} &= -8\pi h f_i \sin 2\pi(hx_i + Lz_i) \sin 2\pi k y_i \\
\frac{\partial A}{\partial z_i} &= -8\pi L f_i \sin 2\pi(hx_i + Lz_i) \cos 2\pi k y_i \\
\frac{\partial B}{\partial z_i} &= -8\pi L f_i \sin 2\pi(hx_i + Lz_i) \sin 2\pi k y_i
\end{aligned} \qquad 17.$$

The derivative with respect to y has been given in equation 12. Weighting factors were computed on the basis of equation 13.

A structure factor calculation with the x, y, and z parameters obtained from the preliminary refinement procedures was made. Readjustment of scale and temperature factors showed no apparent change in any of the four zones, h0L, h1L, h2L, and h3L. Calculation of ΔF^2_{hkl}, $\partial F^2 / \partial q_i$ and weighting factors for each observed non-zero reflection was accomplished in the usual way employing I.B.M. techniques.

After formation of the normal equations, iterative methods were used for their solution. The thirty six equations were solved in about eight hours by the procedure of Gauss and Seidel. The shifts, together with the improved parameters, to four decimal places are listed in Table V.

TABLE V.

RESULTS OF FIRST THREE-DIMENSIONAL LEAST SQUARES

	Parameter	Shift	Corrected Parameters	Probable Errors (A.)
x	C_1	0.0000	0.3410	0.019
	C_2	-.0032	.2868	.021
	C_3	-.0052	.3188	.020
	C_4	-.0005	.3235	.022
	C_5	-.0074	.1526	.021
	N_1	-.0028	.3152	.017
	N_2	.0037	.3537	.018
	O_1	-.0009	.3301	.016
	O_2	.0010	.3910	.017
	O_3	-.0036	.2904	.016
	O_4	-.0004	.0676	.016
	S	-.0004	.0316	.008
z	C_1	-.0010	.0220	.022
	C_2	.0020	.0690	.024
	C_3	-.0049	.2241	.023
	C_4	-.0014	.2956	.022
	C_5	-.0040	.0020	.024
	N_1	-.0017	.1533	.019
	N_2	.0007	.3767	.019
	O_1	-.0033	.4507	.017
	O_2	.0003	.0563	.017
	O_3	.0050	.2200	.016
	O_4	-.0020	.2570	.016
	S	.0004	.4564	.007

TABLE V.(CONT'D.)

	Parameter	Shift	Corrected Parameters	Probable Errors (A.)
y	C_1	.0012	.0229	.037
	C_2	-.0042	.8385	.038
	C_3	.0013	.8490	.037
	C_4	.0006	.0633	.037
	C_5	.0053	.7120	.038
	N_1	.0066	.9903	.030
	N_2	.0036	.8913	.032
	O_1	-.0007	.4170	.028
	O_2	-.0017	.2410	.027
	O_3	-.0067	.6110	.027
	O_4	-.0049	.5918	.027
	S	.0033	.5000	.018

An estimation of the probable errors in the coordinates is also given. These errors were computed from the relation (30):

$$\delta q_i = \left[\frac{\sqrt{w_{hkl}} \Delta F^2_{hkl}}{m - s} \cdot D_{ii}^{-1} \right]^{1/2} \qquad 18.$$

where m is the number of observational equations, s is the number of unknowns and D_{ii}^{-1} is the i^{th} diagonal element of the matrix inverse to the matrix of the normal equation coefficients. Assuming small off diagonal elements:

$$D_{ii}^{-1} \simeq \frac{1}{D_{ii}}$$

Several surprisingly large x and z parameter changes were obtained. These are undoubtedly due to the inclusion

of the reflections from three extra zones in the three-dimensional least squares which were not incorporated in the preliminary refinement.

Structure factors calculated using the corrected parameters showed marked improvement. The R values for the four zones are presented in Table VI. It was noted, however, that a number of rather small observed intensities were in rather poor agreement with calculated values. To emphasize these reflections in the next least squares solution they were given higher weights defined by the equation:

$$\sqrt{w_{hkl}} = 100/F^2_{hkl} \quad \text{for } F^2_{hkl} \geq 16F^2_{hkl\,min.} \qquad 19.$$

$$\sqrt{w_{hkl}} = 2 - F^2_{hkl}/\,100 \quad \text{for } F^2_{hkl} \leq 16F^2_{hkl\,min}$$

TABLE VI.

RELIABILITY FACTORS AFTER FIRST THREE-DIMENSIONAL LEAST SQUARES

Zone	R
h0L	16.4%
h1L	18.6%
h2L	18.6%
h3L	17.6%

The calculation of observational and normal equation coefficients was performed in the routine manner. The coordinate shifts from this second three-dimensional least squares appear in Table VII. The corrected x, y, and z parameters are also listed with their probable errors. As might be

TABLE VII.

RESULTS OF SECOND THREE-DIMENSIONAL LEAST SQUARES

	Parameter	Shift	Corrected Parameters	Probable Errors (A.)
x	C_1	-.0004	.3406	.011
	C_2	.0005	.2873	.012
	C_3	-.0032	.3156	.013
	C_4	-.0003	.3232	.012
	C_5	.0006	.1532	.012
	N_1	-.0013	.3139	.010
	N_2	.0009	.3546	.010
	O_1	.0000	.3301	.009
	O_2	.0002	.3912	.009
	O_3	.0004	.2908	.009
	O_4	-.0013	.0663	.009
	S	.0000	.0316	.004
z	C_1	-.0020	.0200	.014
	C_2	-.0017	.0673	.015
	C_3	-.0022	.2219	.015
	C_4	-.0002	.2954	.014
	C_5	.0012	.0032	.015
	N_1	-.0009	.1524	.010
	N_2	.0010	.3777	.012
	O_1	.0001	.4508	.009
	O_2	-.0010	.0553	.010
	O_3	.0010	.2210	.009
	O_4	-.0010	.2560	.010
	S	-.0001	.4563	.003

TABLE VII (CONT'D.)

	Parameter	Shift	Corrected Parameters	Probable Errors (A.)
y	C_1	-.0020	.0209	.025
	C_2	.0064	.8449	.023
	C_3	.0049	.8539	.025
	C_4	-.0095	.0538	.023
	C_5	-.0030	.7090	.023
	N_1	-.0008	.9895	.022
	N_2	.0049	.8962	.019
	O_1	-.0040	.4130	.015
	O_2	-.0007	.2403	.015
	O_3	.0022	.6132	.015
	O_4	-.0027	.5891	.015
	S	-.0011	.4989	.007

expected, the errors in the y parameters are about twice the x and z errors. This arises from the dependence of the errors on the reciprocals of the D_{ii} terms in the determinant representing the normal equation coefficients. These diagonal elements are roughly proportional to the squares of the lattice translations, x corresponding to a, etc., and thus the D_{ii}'s for the y parameters are much smaller than those for x and z. It should be stated that the errors calculated from equation 18 are standard deviations and must be multiplied by 0.675 to give probable errors.

The corrected parameters of the atoms were used to cal-

culate a final set of structure factors. Hydrogen atoms had not been included in the work up to this point but as this was to be the last calculation they were now introduced. The parameters of the eleven hydrogens in the asymmetric unit were estimated from projections of the structure. The C-H, N-H and O-H distances were all taken as 0.9A., a procedure suggested by various other organic crystal structure determinations (33). The parameters so obtained are given in Table VIII. Two separate sets of structure factors, one with and one without hydrogen atoms, were prepared. It was readily seen that the inclusion of the hydrogen atoms improved the agreement markedly. A small scale factor adjustment in the h3L zone made this agreement still better. Table IX gives a list of intensities computed from these hydrogen-containing structure factors, and also the corrected observed intensities on an absolute scale for about 600 general hkl reflections. The final reliability factors for the four zones are presented in Table X.

TABLE VIII.

Parameters of Hydrogen Atoms

Atom	x	y	z
H_1	0.145	0.130	0.453
H_2	.132	.585	.033
H_3	.349	.712	.098
H_4	.324	.172	.158
H_5	.382	.184	.307
H_6	.245	.127	.280

TABLE VIII (CONT'D.)

Atom	x	y	z
H_7	.443	.867	.405
H_8	.310	.750	.378
H_9	.348	.036	.410
H_{10}	.127	.594	.247
H_{11}	.040	.752	.256

TABLE IX.

FINAL STRUCTURE FACTOR CALCULATION FOR

NN'-DIGLYCYL-CYSTINE

h	k	l	F^2_{obs}	F^2_{calc}
o	o	2	44.0	43.6
		4	79.5	84.6
		6	30.5	31.8
		8	15.1	11.7
		10	36.3	39.5
		12	25.2	26.0
1		o	10.6	11.1
1		2	4.6	1.9
1		4	58.5	59.5
1		6	16.5	18.7
1		10	14.0	15.8
1		14	21.9	25.5
1		16	4.0	4.2
2		o	44.8	46.3
2		2	49.1	48.1
2		4	26.5	22.4
2		10	17.0	15.3
2		14	10.9	9.0
3		o	173.1	173.1
3		2	30.8	33.3
3		4	11.6	12.5
3		6	4.9	2.7
3		8	15.0	17.8
3		10	13.3	12.5
3		12	22.0	25.5
4		o	23.8	24.3
4		2	6.5	2.4
4		4	36.1	40.3

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
4		6	9.1	11.1
4		14	14.6	12.9
5		0	28.4	33.8
5		2	47.1	42.0
5		4	11.0	5.9
5		6	21.0	19.9
5		8	9.5	7.7
5		12	2.6	3.3
6		0	40.3	44.7
6		2	39.5	42.9
6		4	11.4	7.7
6		6	21.9	24.8
6		8	4.1	2.4
6		12	7.0	10.7
7		0	5.0	2.7
7		2	13.0	17.0
7		4	24.8	29.5
7		6	16.1	20.2
8		0	16.0	22.9
8		2	10.2	9.8
8		6	8.8	6.6
8		8	10.3	8.6
9		0	4.0	5.4
9		2	9.5	13.6
9		4	3.1	1.7
9		6	9.5	11.0
11		0	10.0	13.5
1		2-	30.2	27.4
1		4-	38.9	44.2
1		6-	13.8	10.1
1		10-	18.9	19.7
1		12-	17.0	12.7
1		16-	4.5	5.6
2		2-	20.1	22.6
2		4-	4.5	1.6
2		6-	61.0	61.1
2		8-	18.9	17.3
2		10-	28.6	30.6
2		12-	15.2	13.9
2		14-	36.0	38.4
2		16-	5.9	7.4
3		2-	23.0	24.4
3		4-	32.8	29.8
3		6-	47.9	47.0
3		8-	25.3	21.4
3		10-	60.8	63.4

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
3		12-	33.5	30.7
3		14-	5.0	8.9
3		16-	3.3	1.0
3		18-	4.0	6.7
3		20-	3.4	5.0
4		2-	28.6	29.4
4		4-	47.8	48.8
4		6-	47.3	42.7
4		8-	47.0	43.0
4		10-	15.7	14.7
4		12-	14.9	13.5
4		14-	27.0	25.0
4		16-	4.2	4.8
4		18-	2.6	5.8
4		20-	6.1	7.2
5		2-	49.5	54.0
5		4-	30.3	32.7
5		6-	34.1	25.3
5		8-	44.2	48.2
5		10-	21.1	18.6
5		14-	35.0	31.9
5		16-	20.1	16.0
5		18-	7.0	5.7
5		20-	3.5	9.5
6		2-	3.0	4.3
6		6-	2.3	1.3
6		10-	25.4	25.2
6		18-	8.2	10.7
6		20-	2.7	6.1
7		2-	8.2	14.3
7		4-	39.5	44.8
7		6-	35.8	40.6
7		8-	38.7	41.7
7		10-	4.1	7.4
7		12-	9.5	5.1
7		14-	65.3	57.4
7		16-	17.0	12.3
7		18-	20.9	22.1
7		20-	8.6	6.7
8		2-	20.0	29.6
8		4-	21.1	26.8
8		6-	6.0	.3
8		8-	30.0	35.7
8		12-	15.9	19.3

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
8		14-	14.2	9.5
8		16-	20.7	15.3
8		18-	13.0	12.3
8		20-	7.1	6.7
9		2-	22.1	21.4
9		4-	21.9	23.7
9		8-	9.2	10.8
9		12-	8.0	7.6
10		2-	18.0	22.1
10		4-	24.9	28.6
10		6-	31.6	27.7
10		8-	23.9	25.4
10		12-	5.9	7.1
10		14-	44.3	40.0
10		16-	5.3	4.2
10		18-	18.2	17.4
11		4-	12.5	17.1
11		6-	6.0	4.2
11		12-	5.0	7.6
11		16-	7.2	7.1
12		2-	11.0	11.0
12		4-	3.0	6.1
12		10-	4.2	4.6
12		12-	17.7	15.9
12		14-	5.2	9.2
12		16-	9.8	11.9
13		2-	8.0	11.3
13		6-	6.5	12.7
13		10-	4.8	5.3
13		12-	12.9	16.6
13		14-	21.8	20.9
13		18-	11.5	10.7
13		20-	5.2	2.9
14		12-	3.0	7.0
14		16-	5.8	7.3
15		12-	6.7	9.1
15		16-	6.0	9.1
16		10-	3.1	2.3
0	1	3	52.0	57.4
	1	5	38.4	35.1
	1	7	13.7	13.7
	1	9	44.5	42.1
	1	11	20.0	22.7
	1	13	9.1	9.0
	1	15	6.4	9.1
1	1	1	12.9	11.1
1	1	3	31.0	36.5

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
1	1	5	31.3	36.3
1	1	7	28.0	30.9
1	1	9	7.0	5.1
1	1	11	16.5	18.5
1	1	13	10.1	10.9
1	1	15	9.0	9.3
2	1	1	63.8	71.9
2	1	3	22.8	17.7
2	1	5	30.0	37.4
2	1	7	9.0	14.9
2	1	9	7.0	13.0
2	1	11	8.4	12.7
2	1	13	12.2	8.9
3	1	1	68.9	76.6
3	1	3	35.1	40.8
3	1	5	27.0	32.9
3	1	7	12.5	11.3
3	1	9	12.2	13.1
3	1	11	14.5	17.2
4	1	1	28.4	30.0
4	1	3	12.0	9.3
4	1	5	28.0	32.9
4	1	11	7.0	7.8
4	1	13	3.4	5.2
5	1	1	31.6	38.9
5	1	3	15.0	8.8
5	1	5	18.3	23.5
5	1	7	6.0	9.3
5	1	9	4.4	10.4
6	1	1	33.4	33.1
6	1	3	19.8	23.6
6	1	5	21.0	22.7
6	1	7	10.0	10.7
6	1	11	3.0	6.7
7	1	3	9.0	10.3
7	1	5	9.0	9.3
7	1	7	3.0	7.9
7	1	9	2.1	2.9
8	1	1	5.0	8.7
8	1	3	10.3	6.3
8	1	5	9.4	9.0
8	1	7	3.8	4.2
9	1	1	5.6	6.5
9	1	3	4.9	9.3
9	1	5	5.2	7.2
9	1	7	6.0	8.7

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
11	1	1	3.2	5.2
1	1	3-	20.4	21.9
1	1	5-	40.9	42.7
1	1	7-	26.9	24.2
1	1	9-	8.7	9.0
1	1	11-	13.5	14.3
1	1	13-	15.5	13.3
1	1	15-	9.0	8.5
1	1	17-	4.0	5.7
2	1	1-	44.2	46.9
2	1	3-	13.2	13.9
2	1	5-	20.0	20.8
2	1	7-	58.2	55.3
2	1	9-	15.0	16.3
2	1	11-	26.7	22.8
2	1	13-	15.7	17.9
2	1	15-	7.4	8.8
2	1	19-	3.0	4.5
3	1	1-	44.0	48.9
3	1	3-	64.2	66.2
3	1	5-	23.8	21.2
3	1	7-	58.0	52.2
3	1	9-	61.6	60.4
3	1	11-	15.5	12.6
9	1	13-	13.7	14.2
9	1	15-	10.0	6.9
9	1	17-	11.7	11.2
9	1	19-	12.0	9.7
9	1	21-	3.5	3.7
10	1	1-	3.0	5.8
10	1	3-	23.4	29.4
10	1	5-	20.3	22.6
10	1	7-	8.6	9.9
10	1	9-	6.3	6.7
10	1	11-	4.5	8.3
10	1	13-	15.5	13.7
10	1	15-	23.6	24.0
10	1	17-	13.0	11.2
10	1	19-	6.5	5.7
10	1	21-	8.5	5.8
11	1	1-	3.0	9.4
11	1	3-	6.9	6.7
11	1	5-	7.0	6.1
11	1	7-	12.8	11.1

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
11	1	9-	5.0	9.2
11	1	11-	2.9	4.1
11	1	13-	3.6	7.4
11	1	15-	10.3	9.1
11	1	17-	3.8	5.9
11	1	19-	3.3	1.8
11	1	21-	3.0	2.1
12	1	1-	5.8	4.5
12	1	3-	5.0	8.7
12	1	5-	2.4	1.5
12	1	7-	5.5	9.7
12	1	9-	3.5	7.6
12	1	11-	4.0	1.6
12	1	13-	13.8	16.2
12	1	15-	2.5	4.5
12	1	17-	11.7	11.1
12	1	19-	6.0	4.3
13	1	1-	2.6	2.7
13	1	3-	10.0	16.2
13	1	5-	4.1	6.8
13	1	7-	2.9	3.3
13	1	9-	4.0	7.4
13	1	11-	5.2	6.8
13	1	13-	15.2	13.4
13	1	15-	11.8	12.0
13	1	17-	3.9	4.3
13	1	19-	1.3	1.0
14	1	15-	7.5	6.3
15	1	9-	3.5	5.4
15	1	11-	5.5	3.1
15	1	13-	6.0	8.0
15	1	15-	2.5	3.0
0	2	0	49.2	54.2
	2	2	33.7	36.5
	2	4	34.9	37.2
	2	6	3.6	4.7
	2	8	9.8	11.0
	2	10	18.3	17.8
	2	12	14.5	13.9
	2	14	8.0	8.0
	2	16	6.3	5.2
1	2	0	41.2	44.1
1	2	2	33.0	32.9
1	2	4	38.3	34.7
1	2	6	5.0	9.1
1	2	8	24.8	19.9
1	2	10	21.3	22.7

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
1	2	12	14.1	11.9
1	2	14	13.2	12.4
2	2	0	19.2	20.1
2	2	2	29.0	27.8
2	2	4	10.9	11.7
2	2	6	12.6	13.5
2	2	8	7.5	5.4
2	2	12	4.1	2.2
3	2	0	54.1	56.6
3	2	2	45.5	45.8
3	2	4	29.3	29.7
3	2	6	19.8	22.7
3	2	8	8.7	7.7
3	2	10	4.3	6.7
3	2	12	9.0	7.6
3	2	14	2.0	3.7
4	2	0	24.5	22.6
4	2	2	24.0	25.5
4	2	4	21.4	21.8
4	2	6	15.1	12.9
4	2	8	9.7	6.5
4	2	10	5.3	4.4
4	2	12	5.8	5.5
5	2	0	13.2	10.9
5	2	2	23.0	25.1
5	2	4	14.6	12.7
5	2	6	8.0	6.3
5	2	8	7.6	7.7
6	2	0	21.7	24.6
6	2	2	20.0	22.8
6	2	4	12.5	12.8
6	2	6	8.0	11.7
7	2	2	3.0	.7
7	2	4	4.5	7.1
7	2	6	9.0	8.7
8	2	0	8.5	11.2
8	2	2	10.4	10.9
8	2	4	11.4	10.7
8	2	6	7.0	6.5
8	2	8	3.5	5.7
9	2	0	7.0	11.2
9	2	2	8.2	9.9
9	2	4	5.7	6.0
9	2	6	2.4	4.0
1	2	2-	16.5	15.4

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
1	2	4-	16.0	15.4
1	2	6-	45.1	43.0
1	2	8-	17.7	19.8
1	2	10-	22.2	22.1
1	2	12-	9.7	9.1
2	2	2-	31.6	28.1
2	2	4-	36.7	33.5
2	2	6-	10.0	12.1
2	2	8-	27.5	21.1
2	2	10-	27.5	29.4
2	2	12-	15.3	13.1
2	2	14-	18.6	15.9
2	2	16-	4.5	7.5
2	2	18-	2.3	2.1
3	2	2-	16.0	19.3
3	2	4-	26.5	22.5
3	2	6-	7.0	6.4
3	2	8-	15.7	16.2
3	2	10-	15.6	16.4
3	2	12-	7.1	6.9
3	2	14-	5.2	4.6
3	2	18-	6.2	4.1
4	2	2-	18.1	16.8
4	2	4-	3.1	5.0
4	2	6-	32.0	30.3
4	2	8-	22.4	18.8
4	2	10-	22.9	21.0
4	2	12-	14.0	10.1
4	2	14-	14.5	8.7
4	2	16-	12.0	9.1
4	2	18-	10.5	7.7
5	2	2-	5.5	9.1
5	2	4-	25.6	27.4
5	2	6-	26.1	19.8
5	2	8-	32.0	30.5
5	2	10-	19.8	11.8
5	2	12-	3.8	5.1
5	2	14-	14.7	9.6
5	2	16-	6.7	6.2
5	2	18-	3.5	3.5
6	2	2-	15.6	13.1
6	2	4-	22.6	23.1
6	2	6-	15.6	17.4
6	2	8-	40.0	37.7
6	2	10-	26.7	22.4

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
6	2	12-	6.5	7.9
6	2	14-	9.0	6.1
6	2	16-	11.1	7.3
6	2	18-	9.2	6.9
6	2	20-	6.0	5.1
7	2	2-	7.6	7.8
7	2	4-	20.9	19.0
7	2	6-	25.1	23.2
7	2	8-	17.4	17.2
7	2	10-	17.6	15.0
7	2	12-	18.9	13.1
7	2	14-	12.8	8.1
7	2	16-	6.0	5.8
7	2	18-	9.5	8.3
7	2	20-	7.2	5.1
8	2	2-	9.1	8.6
8	2	4-	28.7	28.4
8	2	6-	29.2	23.4
8	2	8-	22.3	24.7
8	2	10-	7.0	4.6
8	2	14-	9.3	10.2
8	2	16-	9.0	6.5
8	2	20-	5.0	2.6
9	2	2-	6.0	8.1
9	2	4-	6.0	10.2
9	2	8-	20.8	20.8
9	2	10-	9.3	9.8
9	2	12-	4.0	3.2
9	2	14-	13.5	10.5
9	2	16-	15.3	10.2
9	2	18-	13.7	9.1
9	2	20-	5.7	4.8
10	2	2-	5.3	8.5
10	2	4-	9.3	10.3
10	2	6-	5.0	6.8
10	2	8-	3.0	7.8
10	2	10-	8.0	6.0
10	2	12-	14.6	11.7
10	2	14-	19.3	15.3
10	2	16-	5.2	6.7
10	2	18-	11.2	7.7
10	2	20-	8.5	4.7
11	2	2-	4.6	6.1
11	2	4-	12.1	11.6
11	2	6-	14.6	11.7

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
11	2	10-	2.3	5.4
11	2	12-	6.4	6.4
11	2	14-	8.5	8.5
11	2	16-	14.2	9.5
11	2	18-	2.1	3.7
11	2	20-	4.4	2.8
12	2	2-	4.3	5.2
12	2	8-	8.0	6.2
12	2	12-	4.5	5.3
12	2	14-	3.3	4.8
12	2	16-	5.0	5.3
13	2	2-	4.0	7.4
13	2	4-	6.9	8.1
13	2	6-	4.5	5.8
13	2	8-	2.0	5.8
13	2	12-	5.3	7.1
13	2	14-	8.5	8.6
14	2	6-	4.4	4.4
14	2	16-	3.0	3.5
0	3	3	12.8	13.2
	3	5	10.9	11.1
	3	7	23.4	22.9
	3	9	11.5	10.4
	3	11	12.0	9.9
	3	13	3.8	4.4
	3	15	3.1	3.0
1	3	1	27.6	27.6
1	3	3	22.0	23.2
1	3	5	14.3	14.2
1	3	7	18.8	18.7
1	3	9	12.3	10.0
1	3	11	10.1	9.6
1	3	13	7.0	6.2
2	3	1	19.4	19.3
2	3	3	19.0	21.7
2	3	5	9.1	10.9
2	3	7	6.6	6.8
2	3	9	3.0	4.0
2	3	11	4.0	5.3
2	3	13	3.7	4.2
3	3	1	20.1	20.5
3	3	3	9.7	10.0
3	3	5	4.8	5.2
3	3	7	10.2	10.7
3	3	9	3.4	4.4

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
3	3	11	3.7	3.6
3	3	13	3.4	3.8
4	3	1	10.2	9.9
4	3	3	12.8	12.9
4	3	5	9.9	11.9
4	3	7	6.5	11.1
4	3	11	3.3	4.3
5	3	1	10.7	11.9
5	3	3	15.2	17.8
5	3	5	9.3	13.1
5	3	7	2.8	2.9
6	3	1	10.3	15.0
6	3	3	9.0	11.6
6	3	5	2.9	4.1
6	3	7	4.0	8.1
7	3	3	3.1	3.4
8	3	3	3.2	6.6
9	3	3	2.8	3.7
9	3	5	2.3	2.6
1	3	1-	22.9	23.8
1	3	3-	14.6	15.0
1	3	5-	8.0	6.8
1	3	7-	10.0	8.3
1	3	9-	7.3	8.3
1	3	11-	6.0	6.3
1	3	13-	3.8	4.0
2	3	1-	14.0	14.1
2	3	3-	10.2	10.7
2	3	5-	8.0	4.1
2	3	7-	29.0	28.9
2	3	9-	21.0	17.5
2	3	11-	13.9	11.8
2	3	13-	9.8	8.0
2	3	15-	2.8	1.8
2	3	17-	2.7	2.0
3	3	1-	19.5	21.7
3	3	3-	10.4	8.0
3	3	7-	16.0	16.6
3	3	5-	22.3	22.0
3	3	9-	18.0	16.9
3	3	11-	15.5	15.6
3	3	13-	5.4	5.2
3	3	15-	3.8	3.8
4	3	1-	10.1	10.2
4	3	3-	20.0	24.0
4	3	5-	14.7	13.8
4	3	7-	13.5	17.9

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
4	3	9-	23.4	21.5
4	3	11-	13.0	10.2
4	3	13-	5.7	5.8
4	3	15-	5.9	2.3
5	3	3-	5.5	5.9
5	3	5-	13.8	15.2
5	3	7-	30.0	30.1
5	3	9-	15.9	13.1
5	3	11-	8.4	6.4
5	3	13-	6.3	4.5
5	3	15-	3.1	2.0
5	3	17-	5.0	3.8
6	3	1-	6.0	5.4
6	3	3-	2.3	2.8
6	3	5-	6.2	8.7
6	3	7-	8.3	9.8
6	3	9-	16.2	15.5
6	3	11-	8.1	8.1
6	3	13-	3.8	3.3
6	3	15-	8.7	8.5
6	3	17-	7.4	6.1
6	3	19-	3.8	4.0
7	3	1-	2.2	1.0
7	3	3-	13.8	15.7
7	3	5-	12.0	13.2
7	3	7-	10.2	14.3
7	3	9-	8.5	8.4
7	3	11-	6.3	6.9
7	3	13-	5.5	7.1
7	3	15-	7.0	5.2
7	3	17-	6.0	5.4
7	3	19-	3.6	3.7
8	3	3-	10.6	12.0
8	3	5-	11.2	13.9
8	3	7-	21.0	21.4
8	3	9-	3.7	5.6
8	3	15-	6.1	5.8
8	3	17-	3.2	3.8
8	3	19-	3.5	2.6
9	3	7-	5.0	5.4
9	3	9-	7.9	10.4
9	3	11-	4.3	4.5
9	3	13-	5.2	5.3
9	3	15-	5.9	6.1
9	3	17-	5.7	4.9

TABLE IX (CONT'D.)

h	k	l	F^2_{obs}	F^2_{calc}
10	3	1-	5.7	5.5
10	3	3-	5.5	6.0
10	3	5-	7.9	9.8
10	3	9-	3.5	4.4
10	3	13-	5.4	6.1
10	3	15-	5.4	5.4
10	3	17-	3.8	4.4
0	4	0	10.7	8.8
1	4		12.6	11.8
2	4		16.3	18.3

TABLE X

FINAL RELIABILITY FACTORS

Zone	R
h0L	13.5%
h1L	14.9%
h2L	14.5%
h3L	12.5%

Undoubtedly the refinement process could have been continued beyond this point but the accuracy of the observed data was not great enough to warrant such a step. It is possible that further least squares treatment would have decreased the probable errors of the y parameters.

C) A discussion of the results of this investigation

By using coordinates from the second three-dimensional least squares Figures VII and VIII were drawn showing i) a projection of the structure on the 010 plane, and ii) a projection of all the atoms between x = 0 and x = 1/2 onto the b-c plane. The numbering system used to indicate atoms is also shown in both figures. Table XI gives the distances between bonded atoms and also an estimation of the error in each such distance. This error was calculated from the equation (36):

$$\text{Probable Error in } R_{ij} = \sqrt{\sum_i Q_i^2} \qquad 20.$$

$$\text{where: } Q_i = \frac{\partial R_{ij}}{\partial q_i} \cdot \delta q_i$$

The bond angles and their probable errors are listed in Table XII. No errors for angles involving hydrogen bonds are given since no probable errors in the X-H bond lengths could be calculated. All these values are in reasonable agreement with the results of previous determinations with the following exceptions.

The hydrogen bonds formed by the amide group are all exceedingly long. In fact, the value of 3.31A. for the $N_1 \cdots O_3$ distance, together with the rather poor $N_1\text{-}H_4 \cdots O_3$ angle, makes the supposition of a hydrogen bond between the atoms a disputable matter. The reason for this abnormal distance is to be found in the van der Waals repulsion of the molecules between which the bond is formed. The molecules in question are lying on top of one another in the b direction as shown in Figure VIII and the sulfur atom of

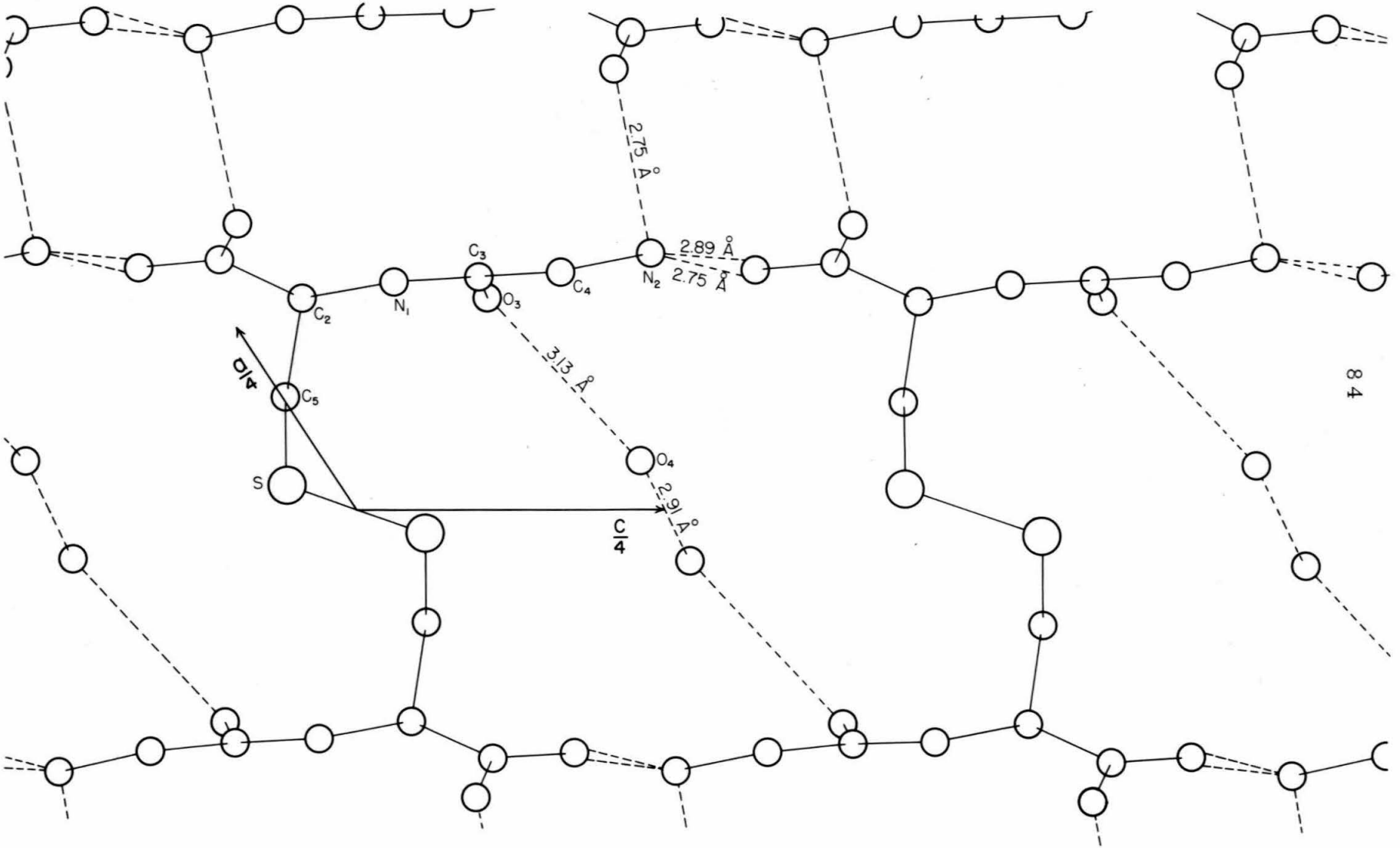


Fig. VII. PROJECTION OF NN'-DIGLYCYL-L-CYSTINE STRUCTURE ON 010

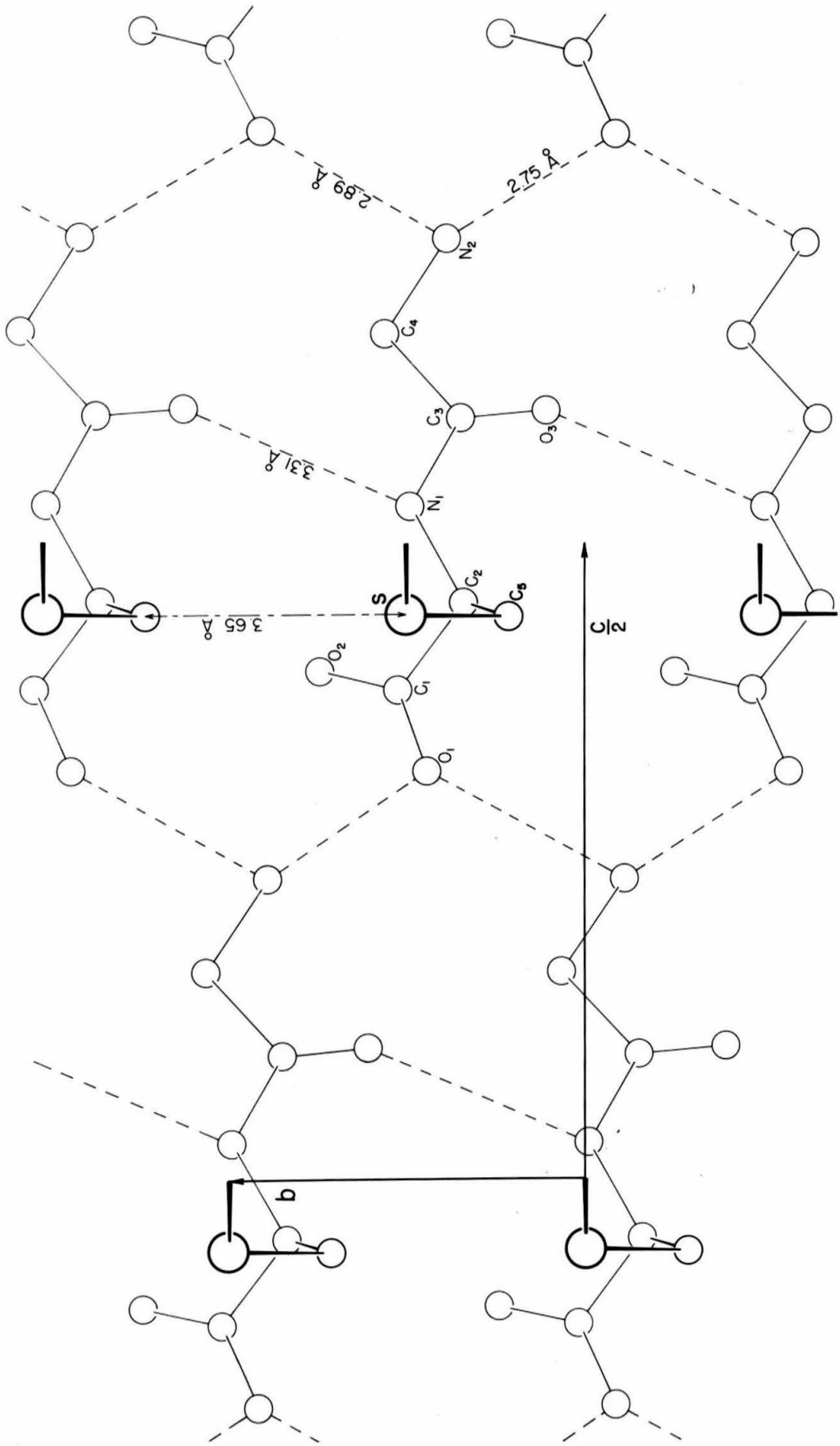


Fig. XIII. PROJECTION OF SECTIONS OF MOLECULES BETWEEN x=0 AND x=$\frac{1}{2}$ ONTO THE b–c PLANE. WATER MOLECULES HAVE BEEN OMITTED.

TABLE XI.

BOND LENGTHS IN NN'-DIGLYCYL-L-CYSTINE

Bond	R_{ij} (A.)	Probable Error (A.)
S-S	2.04	.006
S-C_5	1.87	.020
C_5-C_2	1.52	.019
C_2-C_1	1.55	.025
C_1-O_1	1.24	.025
C_1-O_2	1.21	.021
C_2-N_1	1.48	.025
N_1-C_3	1.35	.026
C_3-O_3	1.21	.022
C_3-C_4	1.55	.024
C_4-N_2	1.46	.022
Hydrogen Bond	R_{ij} (A.)	Probable Error (A.)
N_1--O_3	3.31	.025
N_2--O_1 down	2.75	.023
N_2--O_1 up	2.89	.023
N_2--O_2	2.75	.014
O_4--O_4	2.91	.026
O_4--O_3	3.13	.017

TABLE XII.

BOND ANGLES IN NN'-DIGLYCYL-L-CYSTINE

Angle	(o)	Probable Error (°)
S-S-C_5	103.0	.7
S-C_5-C_2	105.0	.9
C_5-C_2-C_1	117.4	1.6
C_5-C_2-N_1	117.6	1.4
C_1-C_2-N_1	109.4	1.4
O_1-C_1-O_2	126.6	1.6
C_2-C_1-O_1	115.0	1.5
C_2-C_1-O_2	118.5	1.6
C_2-N_1-C_3	121.6	1.6
N_1-C_3-O_3	125.3	1.9
C_4-C_3-O_3	120.6	1.7
N_1-C_3-C_4	113.2	1.5
C_3-C_4-N_2	108.8	1.4

ANGLES INVOLVED IN HYDROGEN BONDS

Angle	(°)
C_4-N_2--O_1 (up)	84.5
C_4-N_2--O_1 (down)	129.8
C_4-N_2--O_2	111.8
C_1-O_1--N_2(2.98A.bond)	143.0
C_1-O_1--N_2(2.75A.bond)	95.6
C_1-O_2--N_2	125.9
C_3-O_3--N_1	144.0
C_2-N_1---O_3	140.1
C_3-N_1---O_3	95.5
C_3-O_3---O_4	106.0
O_3-O_4---O_4	118.2

molecule (1) is in very close contact with the C_5 methylene group of the molecule above. Calculation shows this seperation to be only 3.65A. while the sum of the van der Waals radii of S and CH_2 given by Pauling (9) is 3.85A.* This indicates that the molecules are squeezing together as tightly as possible in an effort to strengthen the N_1-H_4--O_3 bond but it is just not possible to reduce the hydrogen bonded distance below 3.3A.

The reason for the water molecule forming a hydrogen bond to O_3 instead of O_1 has already been given. That this distance is also abnormally long must be explained by the fact that both ways in which this distance might be shortened involve a weakening of other hydrogen bonds. For example, if the oxygen atom of the water molecule were moved toward O_3 the hydrogen bond between O_4 and the water molecule immediately above would be lengthened considerably. Or if O_3 were moved toward O_4 the N_1H_4--O_3 bond would become even longer than it is.

Thus one sees here a case in which two oxygen atoms, O_3 and O_4, each prefer to form two rather weak hydrogen bonds rather than one strong one. This type of behavior supports the belief that number rather than strength of hydrogen bonds may often be the important factor in deciding the

---

* Reporting the structure of 22' Diiododiethyl trisulfide, Donohue (35) suggests that the van der Waals radius of sulfur is too high by about 0.2A. If this is true, the predicted S---CH_2 distance would be just equal to the observed separation. One may assume that the sulfur van der Waals radius is perhaps somewhere between the values given by Pauling and Donohue. In either case it is evident that the two atoms are as close as, or closer than, the repulsion of the outer electrons will allow.
seemingly

structure which will result.

The weak hydrogen bonds formed by O_3 also might account for the rather short C_3-O_3 distance of 1.21A. It is probably true that the strength of the hydrogen bonds formed by such an oxygen atom directly influences the amount of keto-enol resonance in the peptide link. Strong bonds would tend to stabilize the enol configuration giving rise to C-N and C-O bond lengths intermediate to the respective single and double bond distances, while weak bonds would lessen this stabilization, shortening the C-O bond and lengthening the C-N bond. The observed bond lengths in the imide group of diglycylcystine seem to confirm this hypothesis. The C-N and C-O distances, when plotted on a revised curve of bond length vs: bond order (34), suggest that the C_3-N_1 bond has only 1/3 double bond character and the C-O bond 2/3 double bond character as compared to half double bond character for both bonds as found in structures where the hydrogen bonding is stronger.

Still another interesting consequence of these weak hydrogen bonds is the slight non-planarity of the imide group. It has been mentioned in the introduction to this thesis that complete keto-enol resonance would substantially prohibit rotation about the C_3-N_1 bond and ensure the coplanarity of the five atoms C_2, N_1, C_3, O_3, and C_4. If, however, this resonance is not complete one might expect to find small deviations from the rule. This seems to be the case here. A least squares plane for the amide group using all five atoms

involved was calculated and deviations of the atoms from this plane obtained. The results are given in Table XIII. Another least squares plane using only N_1, C_3, O_3, and C_4 was also calculated and the deviations derived therefrom appear in Table XIII under D'. It can be seen that this latter plane is much more suitable if the deviation of atom C_2 is ignored, that is, if a slight rotation about the C_3-N_1 bond is permitted.

TABLE XIII

Atom	D = deviation from Plane 1. (A.)	D' = deviation from Plane 2. (A.)
N_1	0.041	0.022
C_3	.077	.048
O_3	-.018	-.025
C_4	-.055	-.016
C_2	-.051	-.164

This rotation corresponds to only 6°. Pauling has shown that, for complete keto-enol resonance, a rotation of 10° about the C-N bond under consideration should be accompanied by an increase of about 1.0 kcal/mole in the energy of the molecules. Taking into account the smaller amount of resonance here, a rotation of 6° should require about a 0.2-0.3 k cal/mole increase in energy. From Figure VIII it can be seen that the rotation brings O_3 and O_4 closer together without seriously affecting the N_1-H_4--O_3 distance, so presumably the increase in energy caused by rotation is more than accounted for by the stabilization of the O_3--H_{10}-O_4

hydrogen bond.

Hydrogen bonds from the end amine nitrogen atom are all of the usual length although the angles C_4-N_2--X are not as close to 110° as might be desired. The disposition of three hydrogen bonds about N_2 clearly confirms that the molecule exists in the zwitter ion form. The variation in bond length between C_1-O_1 and C_1-O_2 lies in the expected direction since two hydrogen bonds are directed to O_1 and only one to O_2. This configuration would be stabilized if the C_1-O_2 bond had more double bond character than the C_1-O_1 bond, a condition apparently fulfilled in this structure.

Most of the bond angles listed in Table XII are also typical of results previously recorded. Indeed the angles in the diglycine chain bear a marked resemblence to the angles reported for β-diglycine by Hughes (8). The two angles C_5-C_2-C_1 and C_5-C_2-N_1 are, however, substantially larger than the usual tetrahedral angle. No obvious reason for this effect has been found to date although the large C_5-C_2-C_1 angle might be caused by van der Waal's repulsion between the S and O_1 atoms.

A least squares plane was also calculated for the carboxyl group and the deviations of the four atoms O_1, O_2, C_1, and C_2 were found to be all less than 0.01A. The angle between the plane of the carboxyl group and that of the imide group (second plane) was obtained as 19.8°. The amine nitrogen atom is bent out of this latter plane by an angle of 14.8°.

The C_2-C_5 bond makes an angle of about 118° with the plane of the peptide group giving a cis configuration about atom C_2. The C-S-S-C dihedral angle is 101°, some 19° higher than the S-S-S-C angle in diiodo-diethyl-trisulfide as reported by Donohue (35), but very near the value found in S_8. The other bond distances and angles in the bridge, as listed in Tables XI and XII, are not in serious disagreement with the results obtained for other sulfur containing organic crystals.

The separation produced by the bridge between the two diglycine "arms" of the molecule is about 6.3A. The distance between the sulfur atoms and the waters of hydration is 3.75A., 0.50A. above the sum of the van der Waals radii of sulfur and oxygen. The S---C_5 van der Waals contact appears to be the only one which is smaller than usual, the reason for this effect having been discussed above.

The diglycylcystine molecules themselves are bound together in all three directions by reasonably strong hydrogen bonds. If the molecules are visualized as parallelopipeds with two long and one short axis, the overall structure can be thought of as follows: (See Figures VII and VIII.) Parallel to the b-c plane are layers of molecules with their short axes oriented in the b direction, each molecule related to the next in the layer by either the A face centering or the b repeat. Molecules related by the A centering are bound together by hydrogen bonds between the amine group of one and the carboxyl group of the other, while molecules re-

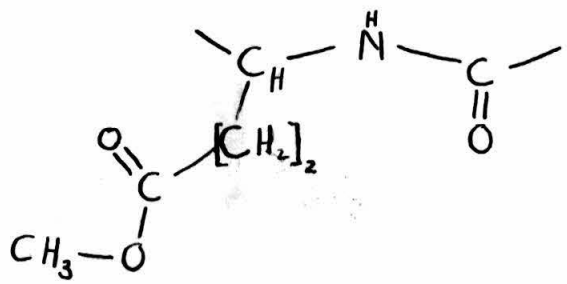
lated by the b repeat are bound by a weak hydrogen bond between their respective imide groups. Also binding each layer together are ropes of water molecules, parallel to b, forming hydrogen bonds to the imide oxygens of the peptide molecules. Successive layers in the a direction are then bound by the N_2-O_2 hydrogen bonds shown in Figure VII. As with all the other peptide structures reported to date, it is clear that the type of packing here is governed by hydrogen bond formation, the molecules forming the maximum number possible.

Related to the study of proteins and polypepties, the significance of this work undoubtedly lies in the elucidation of the structural features of the disulfide bridge set forth above. The non-planarity of the peptide link and the possible smaller contribution of the enol form to the structure of the imide group may also be important in considerations of peptide chain configurations.

IV. A STUDY OF THE STRUCTURE OF POLY-γ-METHYL-L-GLUTAMATE

A) Preparation of oriented fibers of the polypeptide

An account of several previous X-ray investigations on synthetic polypeptides has been given in the introduction. The advantages of studying these substances instead of natural proteins have also been mentioned. One of the most frequently examined peptides has been the polymer of the methyl ester of ~~γ-methyl-L-glutamic~~ acid commonly called poly-γ-methyl-L-glutamate. The chemical formula for a single residue of this compound is:



$C_6NO_3H_9$

Oriented fibers of poly-γ-methyl-L-glutamate give fairly sharp X-ray diffraction diagrams which have been interpreted by Bamford, Hanby, and Happey (21) as indicative of the presence of α_{II} chains and by Pauling and Corey (22), who state that the observed reflections can be described best using an α-helix model. The work to be presented in this section was begun in an effort a) to obtain better diffraction data, b) to interpret this data in the light of the various proposed models of the polypeptide chain and, having selected the model for which best agreement is obtained, c) to derive some quantitative information concerning atomic parameters,

side chain configurations, etc.

Samples of poly-γ-methyl-L-glutamate were prepared by heating solutions of the carboanhydride of the amino acid in various solvents in the presence of a polymerization initiator. The progress of the reactions was followed by observing qualitatively the viscosities of the solutions. Dimethyl formamide, dichloracetic acid and pyridine were found to give the best results as solvents. After allowing the reactions to proceed until what was thought to be an adequate chain length had been attained, the solvents were extracted as far as possible, yielding viscous, jelly-like materials containing probably less than 20% solvent.

These gels were extruded through an ordinary 2cc. hypodermic syringe into n-propanol or mixtures of n-propanol and water. After soaking in the water-alcohol mixture for about an hour the extruded fibers were removed and air dried. It was found that thoroughly dried fibers were fairly tough but not too elastic while fibers still slightly wet could be stretched much more easily.

The stretching of the fibers, during which most of the orientation of the individual peptide chains was doubtlessly accomplished, was performed by carefully pulling the fibers by hand. Extensions in length of about 50% were noted. After this initial stretching the fibers proved to be incapable of further extension. Heating and subsequent annealing of the fibers while under mechanical tension was also employed to improve the orientation, but these procedures were

not tried until a fairly good α fiber had been obtained. The fibers prepared from original solutions in dichlor-acetic acid, dimethyl formamide, and pyridine all showed reasonably good extinctions parallel and perpendicular to the fiber axis when examined under doubly polarized light.

As the fibers from this first preparation were extremely fine, bundles of eight or ten of them, all lying with their long axes parallel, were made up. This was done in order that the X-ray exposure times would not become prohibitively long. An evacuated camera of three centimeter radius was used to record the X-ray diffraction pattern of these fiber bundles. The samples were, of course, stationary. The photographs obtained showed superficial resemblences to the photographs of well oriented films and fibers of poly-γ-methyl-L-glutamate published by Bamford (21) but showed, in addition, spacings which could be attributed to the β, or extended form of the polypeptide chain. In a private communication Bamford stated that this effect has been correlated with a short average length of the chains, the amount of β form present in the samples decreasing markedly with increasing chain length.

No further experiments with the polypeptide were made until about two months after the preparation of the first fibers at which time it was noted that the solution of the peptide in dimethyl formamide had begun to fluoresce with a bright bluish color. None of the other solutions showed this effect. Fibers were extruded from this fluorescent solution,

stretched and photographed with the techniques and apparatus already described. The X-ray diffraction pattern was clearly the same as that reported by Bamford; that is, the fibers gave a well-oriented α diagram with no observable trace of β reflections. It was evident, therefore, that the polymerization reaction had continued in this solution even though the bulk of the solvent had been removed some two months previously. The fluorescence was possibly due to some complex formed between the polypeptide and dimethyl formamide molecules. As it was observed that the solution, as first prepared, fluoresced in the ultra-violet it might be theorized that the dimethyl formamide-peptide complex shows a fluorescence whose wave-length increases with increasing peptide chain length.

No attempt was made to obtain a molecular weight for the polypeptide sample which gave the α fibers, but by analogy with Bamford's materials one may surmise that the average chain must have included about 125 amino acid residues. At this time fibers were also spun from the solutions in pyridine and dichloracetic acid, but they again gave an X-ray pattern containing both α and β reflections. It would thus seem that the velocity of the polymerization reaction is greater in dimethyl formamide than in either of the other solvents employed. Periodically, fibers have been spun from these other solutions but the characteristic β reflections are still present.

It was also observed that, a month after the successful

preparation of α fibers, fibers extruded from the same dimethyl formamide solution could no longer be stretched or oriented by any mechanical means. The fact that, instead of being pliable as before, these new fibers were very brittle suggested that the polymerization had proceeded to such an extent that the intertwined chains in the unoriented fibers were too long to be easily untangled by mechanical stretching. It is therefore apparent that there exists a very discrete range of peptide chain length below which mechanical extension of the fiber produces some β form along with the desired α form and above which the chains are too snarled to permit orientation.

The small batch of α fibers prepared after allowing the polymerization to proceed for two months was now subjected to further treatment in an effort to improve the orientation of the peptide chains. X-ray photographs of the fiber before and after each treatment showed whether or not the desired effect had been obtained. By far the most successful of the experiments in this series was that in which the fiber, under mechanical tension, was placed in an oven at 60-70°C and, after two days at this temperature, was slowly annealed by lowering the temperature about 5° per day. Photographs of a fiber so heated showed a noticeable sharpening of the reflections, particularly in the off-equatorial spectra. A reproduction of the diffraction photograph obtained from this fiber using a helium filled cylindrical camera of ten centimeter radius appears as a frontispiece to this

thesis. (In this picture the fiber axis is oriented parallel to the cylindrical camera axis.)

In the following section, a description of the measurements made on X-ray photographs of these well-oriented poly-γ-methyl-L-glutamate fibers will be presented together with a discussion of their significance as applied to the proposed peptide chain models.

B) A careful measurement of spacings on X-ray photographs of poly-γ-methyl-L-glutamate.

1. Measurements of reflections observed using camera of 3 cm. radius.

Photographs of the annealed polypeptide fiber were taken in a 3 cm. cylindrical camera with the fiber axis parallel to the cylinder axis and at an arbitrary azimuth. A series of photographs from this camera showed that there was little or no change in the diffraction pattern as this azimuth angle was varied, a result to be expected for uni-directional orientation.

Another set of photographs was taken with the fiber axis perpendicular to the camera axis and at several azimuth angles. The films clearly indicated an alteration of the pattern of reflections with angle. Only one reflection appeared to be definitely on the equator of these photographs although several were so close to that line that the fuzziness and broadening characteristic of fiber reflections made it seem, on cursory examination, as though they were equator-

ial spectra too. The solitary observed arc was shown, by careful measurement with a traveling telescope, to have a spacing of 1.472kX. In the introduction it was mentioned that the existence of a reflection from the basal planes of the polypeptide fiber at a spacing equal to the translational rise in the fiber direction per amino acid residue was one of the facts supporting the helical model for poly-γ-methyl-L-glutamate advanced by Pauling and Corey. This helical configuration predicts a rise of about 1.5A. per residue, in good agreement with the value found here.

Moreover, since the α-helix model for this particular polypeptide has eighteen amino acid residues in five turns of the helix, one could predict that the "unit cells" into which the fibers packed might possess either 6_1 or 6_5 screw axes depending on the clockwise or counter-clockwise sense of the helix. To check this possibility, a Weissenberg photograph of the fiber was taken so that the basal planes of the fiber would be in reflecting position. Although the spectra on this film were exceedingly weak and diffuse it was established that the reflection at 1.47kX was indeed the only one from the basal planes. Therefore it could be assumed that the "unit cell" of the structure possessed a true 6_1 or 6_5 screw axis and a pseudo- eighteen fold screw axis. In that case the 1.47kX reflection would be the eighteenth order reflection from planes perpendicular to the fiber axis and would lead to a value of 26.50kX for the repeat in the fiber direction.

Measurements of all the reflections on the 3 cm. pictures were made but with not too great accuracy because of their diffuse nature and their relatively small separations on the film. While they will not be presented here in any detail, the interplanar spacings calculated from these measurements were essentially in agreement with those obtained by Bamford, et al., with the possible exception of the layer line separations which tended to give smaller values for the fiber repeat.

The reflections which Bamford attributes to a second phase in the fiber were also noted. Pauling has shown that these, as well as all the other spectra given by the fiber, can be indexed on the basis of a hexagonal cell with a_o = 11.96A., C_o = 27.5A. The reflections obtained here were also indexed on this basis except that C_o was lowered to about 26.5A. It was also found possible to index the picture completely using Bamford's orthorhombic cell with a quintupled c axis, the lattice translations being a= 10.35A., b = 5.98A., C = 26.5A. It should be noted that 10.35A. is equal to 11.96A. · sin 60° while 5.98A. equals 11.96A.· cos 60° so that the indexing of the photographs using two alternate unit cells may by no means be accidental but may merely reflect the possibility of transforming any hexagonal cell into an end-face centered orthorhombic cell by choosing suitable axes. This relationship between the two cells would require Bamford's a and b axes to be doubled. The extinctions created by the end face centering, however, together with the

special ratio of a and b would make possible the indexing of the diagram using Bamford's halved axes.

The fact that Weissenberg photographs show some sort of six-fold symmetry suggests that Pauling and Corey's hexagonal cell is to be preferred. Therefore throughout the rest of this thesis the discussion will be restricted to this cell.

2. Measurements on photographs taken in a camera of 10 cm. radius.

Being unable to obtain sufficiently accurate spacing measurements from the photographs discussed above, pictures taken in an helium-filled cylindrical camera with a radius of 10 cm. were examined. These photographs, while revealing no new reflections, did give a better resolution of the spectra and permitted their positions on the film to be measured with a fair precision.

Actual distances on the photographs were obtained using a traveling telescope capable of giving readings to a thousandth of a millimeter. The broadness of the arcs and especially their extension in the direction parallel to the fiber axis made the attainment of such accuracy impossible, however. A conservative estimate of the probable errors in these readings would be about $\pm$ 0.03 mm. in directions perpendicular to the fiber axis and $\pm$ 0.07 mm. in directions parallel to it. Film shrinkage was also taken into account, the negatives being scored with a standard five centimeter fiducial marker before and after development. The interplanar spacings

were derived from the film coordinates of the reflections by the usual equations. From the probable errors in the positions of the reflections it was calculated that the probable errors in the spacings were of the order of ± 0.02A. Table XIV gives a list of these spacings, their respective intensities on a qualitative scale and the values of d observed by Bamford for the same reflections.

Note that in every case the spacings observed here are smaller than those found by Bamford, indicating that the interplanar distances are critically dependent on the orienting treatment given the fiber. It could be that the spacings shown by the fibers studied here, being smaller, correspond to a greater degree of orientation than obtained by Bamford.

A calculation was now made to determine whether or not all the observed spacings could be accounted for by the hexagonal cell proposed by Pauling and Corey. It was found that all the reflections might be produced by this cell but with slightly revised lattice translations. Assuming Pauling & Corey's indexing of the hk0 reflections to be correct it was possible to calculate values of a for each such spectrum. A weighted average of these values gave a = 11.58A. while the accurate measurement of layer line spacings on these 10 cm. photographs gave a value of 26.38A. for the C axis. These lengths are significantly lower than the values given by Bamford and Pauling and might be attributed to better orientation as stated above.

From the corrected hexagonal cell dimensions and the in-

dices given by Pauling calculated interplanar spacings could be deduced. These are given in Table XIV. The agreement between calculated and observed data is seen to be exceptionally good.

TABLE XIV.

SPACINGS FOR POLY-γ-METHYL-L-GLUTAMATE

HK·L	d_{obs}(A.)	d'_{obs}(A.)	d_{calc}(A.)	Intensity
10·0	10.03	10.35	10.03	V.V.V.S.
11·0	5.78	5.98	5.78	V.S.
20·0	5.05	5.22	5.03	W.
21·0	3.78	3.89	3.79	S.
30·0	3.34	3.45	3.34	M.
22·0	2.90	3.00	2.90	W.
31·0	2.78	2.87	2.78	W.
20·2	5.30	5.51	5.30	W.
10·3	6.63	----	6.61	W.
10·5	4.64	4.82	4.66	V.V.S.
11·5	3.92	4.05	3.90	V.S.
20·5	3.62	3.75	3.63	W.
11·6	3.47	----	3.49	V.W.
10·8	3.14	----	3.13	W.
11·8	2.86	----	2.86	< M.
20·8	2.75	----	2.75	< M.
21·8	2.49	----	2.49	W.
00.18	1.47	----	1.47	M.

hexagonal cell: $\underline{a}$ = 11.58A., $\underline{c}$ = 26.38A.

d_{obs} -- spacings observed in this investigation.

d'_{obs} -- spacings observed by Bamford.

The indices assigned to the reflections show no systematic absences except in the 00·L zone for which no order below 00·18 is observed. This simple extinction is consistant only with either a $C6_5$ or $C6_1$ space group with one molecule per cell. It may be somewhat facetious to speak about a "space group" for a semi-crystalline substance but, as will be apparent in the next section, the structure factor equations for these space groups give calculated intensities in reasonably good agreement with the observed, so that the assignment is perhaps significant.

It is also interesting that the 00·6 and 00·12 reflections, permitted by either of the space groups mentioned, are not present. This suggests that the symmetry element in this structure is a special kind of six-fold screw axis, having eighteen-fold symmetry. The importance of this observation as related to the proposed models of the polypeptide chain will be discussed later.

C) Calculations of X-ray diffraction intensities for the α-helix model of poly-γ-methyl-L-glutamate.

As in any structure determination from X-ray data, the designation of lattice parameters for fibrous poly-γ-methyl-L-glutamate did not finally settle the structure of the molecule. Some indirect evidence in favor of a helical model could be obtained from these distances, but it remained necessary to compute the intensities of the scattered X-ray beam from a fiber composed of molecules having this configuration

and to compare the calculated values with the observed intensities before any definite decision could be made.

To be complete, any such calculation of intensities would have to be made for each of the models proposed for poly-γ-methyl-L-glutamate molecules. At the date of this writing, this would entail calculations for Bamford's α_{II} chain and Pauling and Corey's 3.6 residue α-helix. However, as Bamford in his report of the α_{II} model does not even indirectly suggest how this molecule is to be oriented in the unit cell nor what kind of packing is to be expected, intensity calculations for this structure would be a rather prolonged task. Moreover, as has been shown in the previous pages, the unit cell in which Bamford proposed to place the α_{II} chains has been radically varied in one direction. It is not impossible that the α_{II} model could be altered in such a way as to require a repeat in the chain direction five times as long as that originally proposed, but as such a change has not been suggested it seems pointless to carry out intensity calculations with the α_{II} model as it stands. Therefore it was determined to make the calculations for the 3.6 residue α-helix alone and from the agreement, or lack of agreement, with observed intensities to decide on its possible validity.

The major features of the 3.6 residue helix have been outlined in the introduction. It should be noted that, of the ten atoms constituting a methyl glutamate residue, excluding hydrogen, only four lie on the helix itself, the other six constituting a side chain. In presenting this model, Pauling

and Corey discussed in detail the configuration of the atoms forming the helix but did not propose a model for the side chain groups. The following table gives the parameters of the former atoms, together with two sets of possible parameters for the β carbon atoms (the side chain atom bonded to the helix), in cylindrical coordinates. The spiral nature of the model enables one to obtain the coordinates of the atoms in any other residue in the molecule by applying successive rotations of 360°/3.6 or 100° in ϕ and translations of 1.47A. in z to the parameters given for the first residue.

TABLE XV.

ATOMIC COORDINATES FOR THE 18-RESIDUE 5 TURN HELIX ρ, z in A.

	Atom	ρ	ϕ	z
	C	2.29	0.0°	0.00
	N	1.59	27.8°	0.88
	C'	1.61	73.8°	0.43
	O	1.74	82.0°	-0.76
(set 1)	BC	3.34	-17.6°	0.76
(set 2)	BC	3.34	17.6°	-0.76

The bond distances, angles, and hydrogen bond distances to which these parameters correspond have been discussed by Pauling and Corey (22) in their original publication. In passing, it might be stated thatall these values conform with the usual results of structural determinations on simple amino acids and peptides.

From these coordinates two possible methods of structure factor calculation are available. The first of these involves the direct use of the structure factor formulas perinent to the space group of the polypeptide. Assuming a counter-clockwise spiral, that is one in which successive residues 100° apart in ϕ have increasing z parameters, the space group $C6_5$ must be assigned. The structure factor equations for this space group are:

$$A_i = 2f_i \{\cos 2\pi(hx_i + ky_i + L/4)\cos 2\pi(Lz_i - L/4)$$
$$\cos 2\pi(kx_i + iy_i - L/4)\cos 2\pi(Lz_i - L/12)$$
$$\cos 2\pi(ix_i + hy_i + L/4)\cos 2\pi(Lz_i + L/12)\}$$

$$B_i = 2f_i \{\cos 2\pi(hx_i + ky_i + L/4)\sin 2\pi(Lz_i - L/4) \quad 21.$$
$$\cos 2\pi(kx_i + iy_i - L/4)\sin 2\pi(Lz_i - L/12)$$
$$\cos 2\pi(ix_i + hy_i + L/4)\sin 2\pi(Lz_i + L/12)\}$$

Parameters of all the atoms in a single repeat distance would have to be used in such a calculation if accurate results were to be obtained. However, these were not available for five side chain atoms per residue or 72 atoms per repeat so that any calculation could only employ parameters for 50% of the heavy atoms. It has been suggested by Pauling that the omission of these latter atoms may not affect the results too seriously since their contribution to the scattered intensity would be expected to be less because of greater mutual interference. It may even be true that the positions of the side chains are not fixed by the "crystal symmetry" but are random. This could be the case if the side

chain groups were but little affected by the orienting procedures applied to the fibers. The fact that the atoms of these groups could not be bound in any way by hydrogen bonds but are presumably restricted only by van der Waal's forces tends to give support to this possibility. The net result of such disorientation in the side chains would be, of course, that they would contribute little or nothing to the over-all scattering, the helix atoms and the β carbon atoms accounting for the entire effect. These considerations therefore encouraged calculations with the five atoms whose parameters could be fixed.

The second method of structure factor calculation makes use of the cylindrical nature of the helical model and involves the expansion of the rigorous structure factor equation in terms of Bessel functions of integral order. Pauling and Corey (22) state that for the equatorial, or HK·O spacings the structure factor is given approximately by:

$$F = \sum_i f_i J_0(4\pi \rho_i \sin\theta/\lambda) \qquad 22.$$

where J_0 is the Bessel function of zero order, ρ_i is radius of the i^{th} atom from the center of the helix, and f_i is the atomic scattering factor for the i^{th} atom at the particular value of $\sin\theta/\lambda$. The advantage of this formulation of the structure factor over that of equation 21 lies in the fact that only the atoms of one residue need be considered thus shortening the computation time greatly.

Since the appearance of Pauling and Corey's paper, Vand (37 and Cochran (38) have developed equations giving the

scattered intensity as a function of $\sin\theta/\lambda$ for other layer lines of a fiber pattern from a helical model. Their actual derivations have yet to appear, but should proceed somewhat along the following lines:

Let the first atom of the spiral under consideration lie at r, o, o in cylindrical coordinates. With cylindrical coordinates α, ζ, and ξ in reciprocal space and np, r and β_n in real space, the scattered amplitude is given by:

$$F_i = f_i \sum_{n=1}^{N} e^{2\pi i [np\cdot\zeta + \beta_n \cdot \xi]} \qquad 23.$$

In this equation p is the rise in the fiber direction per residue, N is the total number of atoms of kind i per repeat distance, and β_n equals np/P cycles where P is the pitch of the helix or the rise per turn in the fiber direction.

$$F_i = f_i \sum_{n} e^{2\pi i [np\zeta + r\xi \cos 2\pi(\alpha - \beta_n)]} \qquad 24.$$

ζ equals L/c in reciprocal space because of the c repeat. Therefore:

$$F_i = f_i \sum_{n} e^{2\pi i [np L/c + r\xi \cos 2\pi(\alpha - \beta_n)]} \qquad 25.$$

If the factor $e^{2\pi i [r\xi \cos 2\pi(\alpha-\beta_n)]}$ is now expanded in a series of Bessel functions one obtains:

$$F_i = f_i \sum_{n} e^{2\pi i np L/c} \sum_{-\infty}^{+\infty} J_k(2\pi r\xi) \cdot e^{2\pi i k(\alpha-\beta_n) + ik\pi/2} \qquad 26.$$

where k is an integer. This can be written as

$$F_i = \sum_{-\infty}^{+\infty} A_k J_k(2\pi r\xi) \qquad 27.$$

where:

$$A_k = f_i \sum_m e^{2\pi i [mpL/c - Rmp/P]} \cdot e^{2\pi i k (\alpha + \frac{1}{4})} \qquad 27.$$

since $\beta_m = np/P$. A_k can be large only when the frequency of the exponential is an integer since we vary n and if $pL/c - kp/P$ is not an integer all values + and - will occur and add to zero. Therefore only Bessel functions of order k such that

$$pL/c - kp/P = m \text{ (an integer)} \qquad 28.$$

are considered; with this restriction

$$A_k = f_i N e^{2\pi i k(\alpha + \frac{1}{4})} \qquad 29.$$

Because of the nature of the Bessel function Cochran (38) states that the summation of equation 27 need not be extended over all values of k, but that only terms with $k \leqslant 4$ should give significant contributions to the scattered wave. He also proceeds to derive from equation 28 the lowest values of k for L ranging from 0 to 28 and finds that the layer lines shown by oriented poly-γ-methyl-L-glutamate are just those for which values of $k \leqslant 4$ can be deduced.

Some features of the above derivation should be noted. Firstly, the result obtained in equation 27 applies to only one set of atoms, say the α carbon atoms of the helix, and that set is also unique in that its first member has the coordinates r, o, o. To obtain the scattered amplitude for the entire α-helix one must add terms like 27 for each set of atoms in the molecule and the terms must be multiplied by a phase factor $e^{-2\pi i k (\phi_j - \frac{m}{R} \cdot z_j/P)}$ to account for the fact that

the i^{th} set of atoms does not have its first member at a point r_i, o, o but at a point r_j, ϕ_j , z_j. In this factor ϕ_{oj} is defined as:

$$\phi_{oj} = \phi_j - z_j/\rho \qquad 30.$$

Equipped with equations 21 and a form of 27, calculations of the intensities as a function of sin θ/λ for L = 0 and 5 were made. Only five atoms were included these being the helical chain atoms and the β carbon atoms although two unsuccessful trial models of the side chain group were brought into the zero layer calculation. Table XV lists two possible sets of β carbon positions and each of these was included in a separate structure factor computation.

Before presenting the results of these calculations it would be well to discuss another feature of the helical model of Pauling and Corey, namely the unknown rotational parameter of the helix about its own axis. By writing down the parameters of all eighteen of a given class of atoms in the molecule, in cylindrical coordinates, it can be seen that if the entire helix is rotated by twenty degrees about its unique axis the resultant parameters will be such that an atom somewhere in the spiral will be brought to the same value of ϕ which the first atom of the spiral had before the rotation process. Since the peptide chain is for all practical purposes continuous and since the structure obviously has no centers of symmetry, this means that the spiral obtained by a twenty degree rotation is identical with the original spiral. For rotations of between zero and twenty degrees, however, the config-

urations are not the same and would be expected to lead to slightly different values of calculated intensities. In an effort to determine this rotational parameter, intensities for the fifth layer line and for a few equatorial reflections were calculated for three separate angles of rotation of the helix, one with the angles ϕ given in Table XV corresponding to a zero rotation, the second with angles of $\phi + 5^{\circ}$ and the third with angles of $\phi + 10^{\circ}$.

As Pauling and Corey have already presented the zero-order Bessel function curve which approximates the scattering on the zero layer line it was not necessary to repeat this calculation. Instead equations 21 were employed to calculate A (B = 0) for the HK·0 reflections with $\sin\theta \leqslant 0.5$. The results of this calculation for the unrotated helix are given in Table XVI along with the values of A for a few reflections using a helix rotated by 10°. Comparison of the calculated structure factors for both models shows so little difference between the two that no determination of the rotational parameter would be possible even if good single crystal intensity data were available.

The values of A for the unrotated helix were subjected to a little further refinement in order to make comparison with observed intensities easier. First the atomic form factors used in the computation contained no temperature factor and, as the photographs showed what seemed to be a rather sizable decrease in intensity with increasing $\sin\theta$, the application of a relatively large temperature factor to the

TABLE XVI.

CALCULATED INTENSITIES FOR POLY-γ-METHYL-L-GLUTAMATE

HK·L	A	A'	A*	A*·m	F^2_{obs}	A(10° rotation)
10·0	1917	1693	6090	6090	V.V.V.S.	1918
11·0	652	449	1217	1217	V.S.	661
20·0	242	147	368	368	W.	231
12·0	-444	-186	-400	-800	S.	-446
30·0	-615	-202	-413	-413	M.	-616
22·0	-643	-145	-267	-267	W.	
13·0	-619	-123	-221	-442	W.	
40·0	-488	-67	-113	-113	--	-489
23·0	-330	-31	-49	-98	--	
14·0	-229	-17	-28	-56	--	
50·0	-49	-2	-3	-3	--	
33·0	21	1	1	1	--	
24·0	50	2	3	6	--	
15·0	125	3	4	8	--	
60·0	196	2	3	3	--	
34·0	202	2	3	6	--	225
25·0	209	2	2	4	--	

$$A' = A \cdot e^{-50 \sin^2\theta/\lambda^2}$$

$$A^* = A' \text{ (Lorentz \& polarization factor)}^{-\frac{1}{2}}$$

values of A was made necessary. The value of B in the exponential $e^{-B\sin^2\theta/\lambda^2}$ was chosen as 50 since this factor brought any calculated intensity at $\sin\theta = 0.5$ to zero, a condition warranted by the character of the photographs. Secondly, because of the impossibility of obtaining observed intensities

of other than qualitative accuracy an inverse Lorentz and polarization factor was applied to the calculated structure factors. Lastly, multiplicity factors caused by the superposition of intensities from several planes in the fiber on the photographs were multiplied into the corrected values of A. The final structure factors are given in the fifth column of Table XVI. Comparison with observed intensities shows good over-all agreement except for the 30·0 reflection which seems to calculate low.

Despite the existence of two sets of β carbon atoms only one value of A appears in Table XVI. This is due to the fact that for the HK·0 reflections it was found that both possibilities lead to very nearly the same calculated structure factors. This is undoubtedly caused by the overlapping, or near overlapping, of atomic coordinates which as far as the HK·0 reflections are concerned, lie on circles at $z = 0$.

Figure IX shows a plot of the results of this calculation on the same scale as the Bessel function curve for equatorial scattering derived by Pauling and Corey (22). It is interesting to note the good agreement between the two calculations, indicating that the Bessel function approximation is indeed adequate for the representation of the equatorial intensities.

The calculation of intensity vs. sin Θ for the fifth layer line of the diffraction pattern from poly-γ-methyl-L-glutamate involves the use of first order Bessel functions

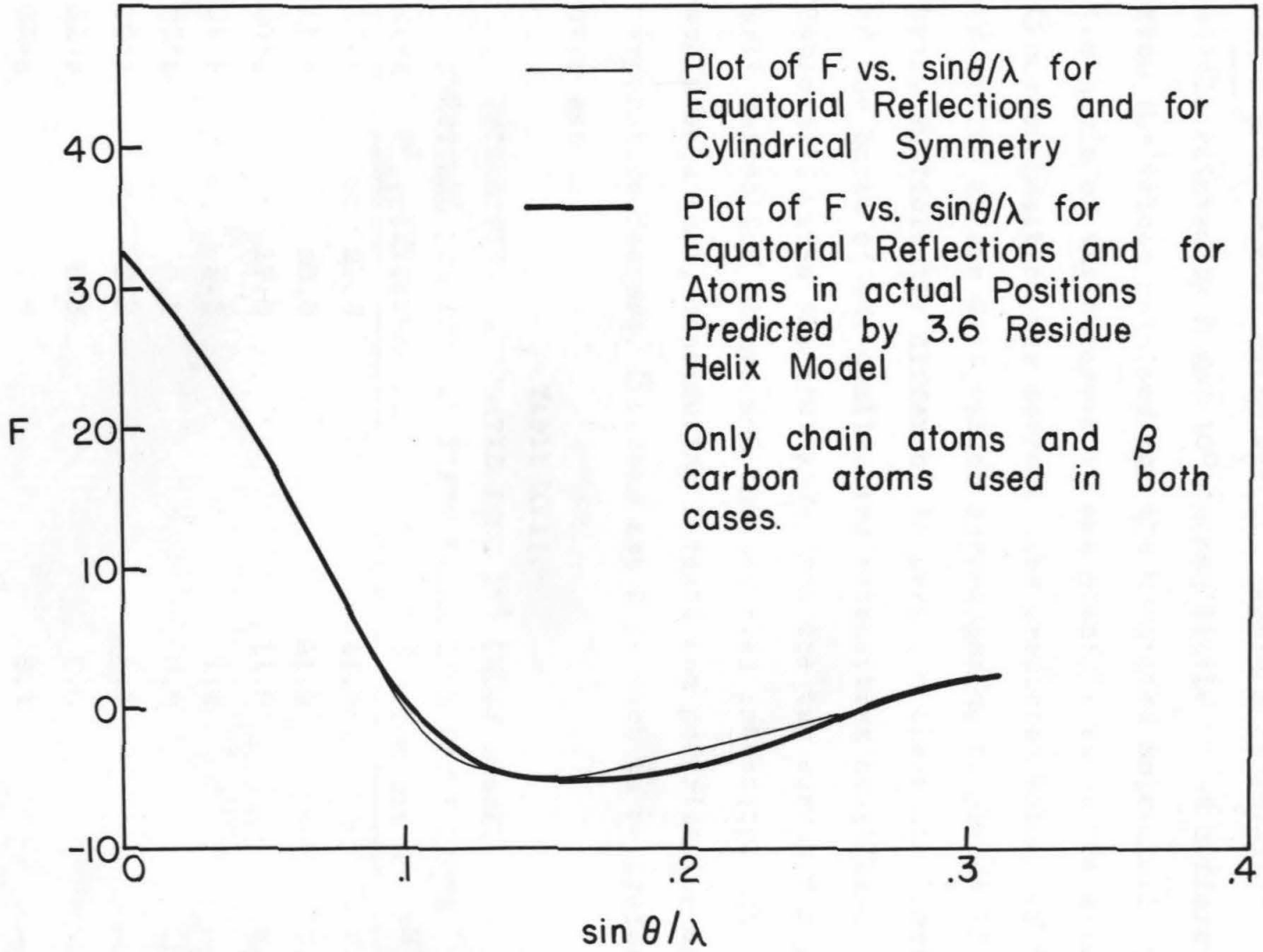

Plot of F vs. sinθ/λ for Equatorial Reflections and for Cylindrical Symmetry
Plot of F vs. sinθ/λ for Equatorial Reflections and for Atoms in actual Positions Predicted by 3.6 Residue Helix Model
Only chain atoms and β carbon atoms used in both cases.
F
40
30
20
10
0
-10
0
.1
.2
.3
.4
sin θ/λ

Fig. IX.

in equation 27. Such a computation was made and the results are plotted in Figure X for the model containing β carbon set 1 and for that using set 2. (Again a calculation for models rotated by 5 and 10° showed little or no differences from the values obtained for the unrotated molecule.) On the basis of these curves it was possible to decide which β carbon position was correct, the predicted values of intensities at sin θ/λ values corresponding to planes HK·5 being sufficiently different to permit a clear distinction on the basis of the qualitative intensities available. Table XVII gives the ordinates from the two curves for several reflection planes and the observed intensity. It would seem that, considering Lorentz and polarization and temperature factors, β carbon set 2 is much to be preferred over set 1.

TABLE XVII.

INTENSITIES CALCULATED FROM 1st ORDER BESSEL FUNCTIONS FOR SEVERAL FIFTH LAYER LINE REFLECTIONS

HK·L	F^2_{calc} (β carbon set 1)	F^2_{calc} (β carbon set 2)	F^2_{obs}
10·5	21.0	41.8	V.V.S.
11·5	20.0	21.9	V.S.
20·5	17.6	11.9	W.
21·5	8.5	1.4	--
22·5	0.1	0.0	--
30·5	2.2	0.1	--
31·5	0.3	0.0	--
32·5	2.5	0.1	--
40·5	3.1	0.1	--
41·5	2.0	0.9	--

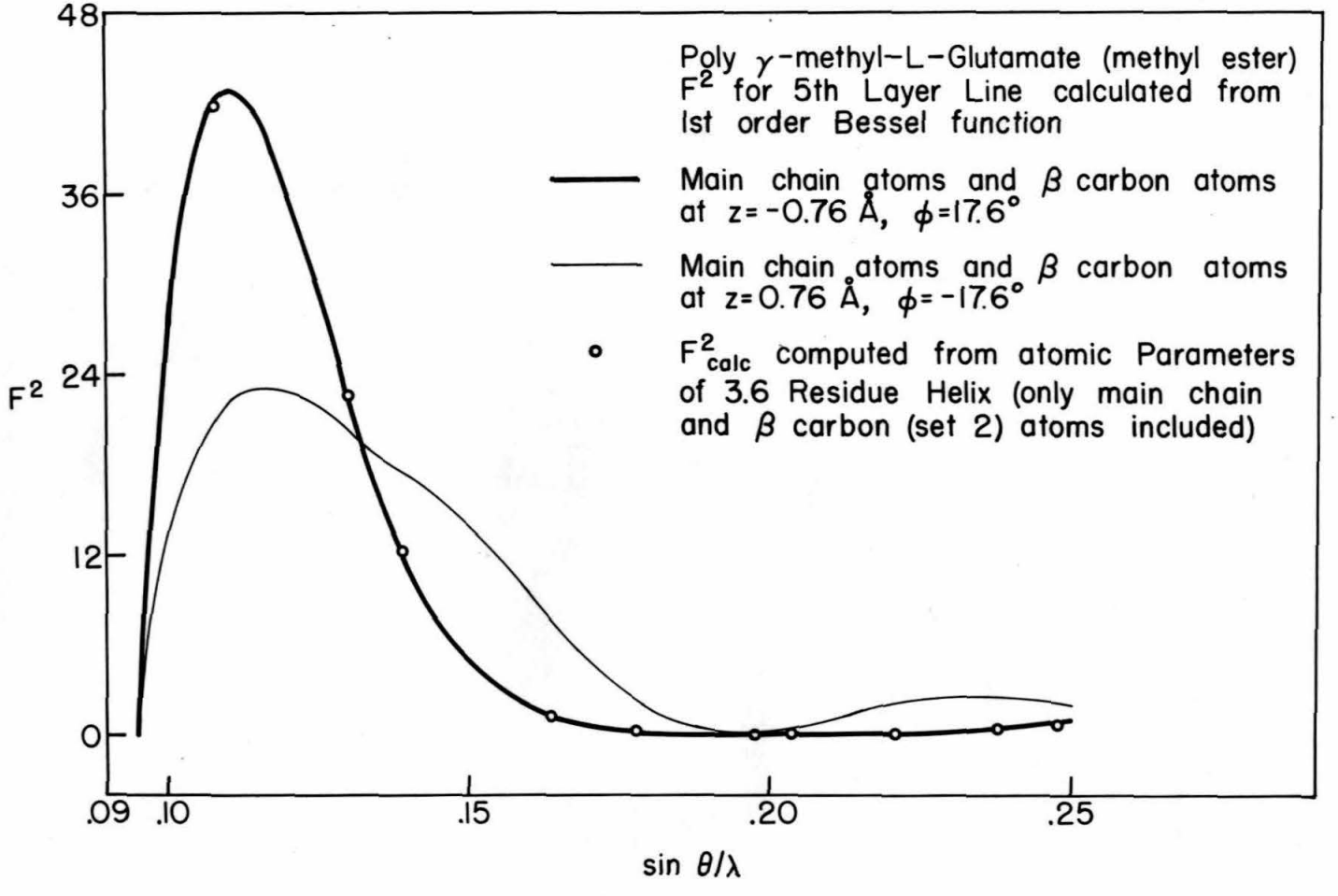

48
36
24
12
0
F^2
.09
.10
.15
.20
.25
sin θ/λ
Poly γ-methyl-L-Glutamate (methyl ester)
F^2 for 5th Layer Line calculated from
1st order Bessel function
Main chain atoms and β carbon atoms
at z=-0.76 Å, ϕ=17.6°
Main chain atoms and β carbon atoms
at z=0.76 Å, ϕ=-17.6°
F^2_{calc} computed from atomic Parameters
of 3.6 Residue Helix (only main chain
and β carbon (set 2) atoms included)

Fig. X.

As a check on this calculation, and in order to determine the correctness of the Bessel function expansion for upper layer lines, a structure factor calculation for fifth layer line planes with $\sin \Theta/\lambda \leq 0.25$ was made using equation 21. The values of A and B obtained are listed in Table XVIII while $A^2 + B^2$, on an altered scale is plotted on the Bessel function curve in Figure X. The agreement is seen to be excellent and completely confirms the validity of the Bessel function formulae.

TABLE XVIII.

STRUCTURE FACTORS CALCULATED FROM EQUATION 21 FOR SEVERAL FIFTH LAYER LINE REFLECTIONS

HK·L	A	B
10·5	200	688
11·5	-183	491
20·5	38	383
21·5	-55	109
22·5	11	15
30·5	-3	48
31·5	19	5
32·5	-5	-58
40·5	7	-23
41·5	-38	-78

In connection with the structure factor calculations employing equation 21 it should be noted that the values for only a few planes in each zone are given. For example, in the HK·5 zone the $10\cdot\overline{5}$ reflection is not listed even though,

according to equations 21, it may have an entirely different structure factor than the 10·5 reflection and even though one would expect to find the two superimposed on a fiber diffraction diagram. However, the unusual 18-fold symmetry of the peptide chains and their helical nature in Pauling and Corey's model causes such planes to give the same calculated structure factors. Therefore only one of them is listed in the table. The same argument holds also for the 12.0 and 21·0 for example which are not crystallographically equivalent but which do give the same calculated intensities.

Further calculations for other layer lines of the poly-γ-methyl-L-glutamate diagram are also being prepared, but as yet no results are available. Owing to the success of the Bessel function expansions in describing the scattering from the helices and to the shorter times required for their calculation the rather tedius computations for equation 21 are being abandoned in this latest work.

D) A critical evaluation of the results

The purpose of this final section is to examine the results reported in the previous pages in terms of the α_{II} and α-helix models of the polypeptide chain. The main comparisons will be made for the α-helix configuration since the structure factor calculations were made for that model alone, but some deductions about the possible correctness of the α_{II} chain from simple spacing data can be given.

First from the nature of the diffraction pattern from

highly oriented fibers of poly-γ-methyl-L-glutamate, it is possible to state positively that the repeat in the fiber direction cannot be as short as supposed by Bamford. If the reflections which are here reported as second, third and eighth layer line spectra were really produced by a secondary phase, one would expect them to vary in shape and intensity from fiber to fiber depending on the orienting treatment and on the manner in which the fibers were prepared. That this is not so is shown by the fact that photographs from a number of poly-γ-methyl-L-glutamate fibers showed no such variation. Therefore it must be assumed that these reflections are truly a part of the polypeptide pattern, and thus require the quintupled c axis repeat.

The elongated c axis is, of course, more easily explained by the helical model than by the flat chain of the α_{II} model which, as originally described, repeats every 5.4A. in the chain direction.

Further evidence in support of the α-helix is obtained in the Weissenberg photographs of the fibers which show as the only order of 00·L present ~~in~~ the 00·18. This is extremely significant in that it not only shows that a six-fold screw axis is present in the cell and, since there is but one molecule per cell, that the molecule must possess a six-fold screw axis, but also indicates that the screw axis is actually eighteen-fold, a symmetry impossible for the unit cell but certainly possible for the polypeptide molecule. The α_{II} chain by no stretch of the imagination could be expected to

have such a symmetry element. The 3.6 residue α-helix, however, has exactly the configuration which would produce such an effect. The eighteen γ-methyl-Lglutamate residues per repeat of this helix are so arranged that the first observable order in the 00.L zone would be the 00·18. Another interesting calculation tending to confirm Pauling and Corey's model can be made from the interplanar spacing of the 00·18 planes and the fifth layer line spacing. If the former quantity is identified with the translational rise in the fiber direction per residue, and the latter with the rise per turn of the helix in the same direction, the ratio should give the number of residues per turn. The ratio of the two observed distances is very nearly 1:3.6 in excellent agreement with the value predicted from the α-helix.

From his derivation of the Bessel function representation of the scattering from a periodic fiber, Cochran has also obtained powerful evidence in support of the 3.6 residue helix model for poly-γ-methyl-L-glutamate. The fact that those layer lines given by the oriented fibers are the only ones for which Bessel functions of order lower than four can be derived from a spiral model seems hardly apt to be mere accident.

In discussing the results of the actual calculation of structure factors for various poly-γ-methyl-L-glutamate reflections one is tempted to see only the good agreement with observed data and to overlook the rather serious approximations involved. The most glaring of these is, naturally, the

omission of half the atoms in the structure from the calculations. The question to be answered here is whether or not the agreement would be as good if the side chain atoms were included in some arbitrary allowed configuration. Pauling and Corey have decided, from the characteristics of the Bessel functions, that there should be little change in the calculated intensities if the side chain atoms are included. An attempt to find out exactly how much these quantities might change was made in this investigation. Two of the many possible side chain configurations were proposed and their parameters included in a calculation of the equatorial or zero layer line intensities using equations 21. It was observed that the resultant structure factors differed by as much as 20-30% from those calculated without the side chain atoms. For one of the models the agreement with observed intensities was made poorer while for the other the over-all agreement was essentially unchanged. With good single crystal intensity data it would be a fairly easy task to find that side chain configuration for which the agreement was best, but with the present data it would seem to be a hopeless undertaking. In fact, as has been stated in the previous section, one cannot even be sure that the fibers under investigation are sufficiently oriented so that the side chains as well as the helices occupy unique positions in the unit cells.

Another difficulty partly caused by the lack of good intensity data is the inability to fix the rotational parameter

of the helix about its axis. Very small differences do exist between structure factors calculated for two helix orientations but the qualitative intensities from the fiber diagrams could never distinguish between them. The large temperature factors needed to explain the rapid falling off of intensities on the fiber photographs makes the problem still more difficult. In this connection, one could hope that part of this apparent temperature effect is due to disorientation and that more nearly crystalline fibers would show larger intensity differences.

However, the above criticisms do not detract from the general evidence which the structure factor calculations give in favor of the helical model. It seems improbable that if the proposed structure were incorrect the calculated intensities would show such good agreement with the observed, even allowing for differences caused by omission of the side chain atoms. No similar calculations for any of the other polypeptide chain models has been made as yet but it does not seem that any of them would give as good agreement as those for the 3.6 residue helix. The results of this investigation, therefore, strongly support the helical model of Pauling and Corey for the "backbone" structure of poly-γ-methyl-L-glutamate.

The significance of this model in the field of protein structure has been stated many times and by many authors. The general notion of a spiral configuration of the polypeptide chain is a very satisfactory one for the structural

chemist since it alone of all the models proposed for this chain permits strong hydrogen bonds between strictly planar amide groups to be formed. Moreover the packing of helical molecules is likely to lead to the high degree of symmetry shown by many crystalline polypeptides, natural and synthetic.

Several attempts to explain the structures of other polypeptides in terms of α-helices have also been made, as noted in the introduction. Of these substances only collagen and keratin have been subjected to further research in effort to prove or disprove the helical model. So far the investigations have tended to show that the proposed models for these compounds need some revision before they will give satisfactory agreement with observed data. It seems likely, however, that the models eventually evolved for these and other substances will retain many features of the original helical configurations.

V.

APPENDIX I.

THE ATTEMPTED APPLICATION OF BANERJEE'S EQUATION TO THE SOLUTION OF THE α-TRIGLYCINE STRUCTURE

The problem of the determination of the phases of X-ray reflections from crystalline solids using only observed data was first investigated by Banerjee (39) who derived some simple relations between the structure factors in a given zone of reflections from a centrosymmetric crystal. Since that time several refinements of the original methods developed by Banerjee have been made (40), and a few applications to actual structure problems have been attempted. The structure of α-triglycine seemed to offer a good opportunity to apply the original Banerjee equations since the long c axis produces many orders of L on Weissenberg photographs and since the crystal is centrosymmetric.

To derive the Banerjee equations for the special case of the hoo reflections, one may begin with the structure factor relation:

$$F(hoo) = \sum_{1}^{m} f_r \, e^{2\pi i x_r \cdot h/a} \qquad (1)$$

If the crystal contains atoms of only one kind,

$$F(hoo)/f_r = S_h = \sum_{r} \beta_r^{\,h} \qquad (2)$$

where:

$$\beta_r = e^{2\pi i x_r/a}$$

If the n values of β_r are the roots of the equation

$$\prod_{r} (x - \beta_r) = x^n + A_1 x^{n-1} + \cdots + A_n = 0 \qquad (3)$$

then, by Newton's relation:

$$
\begin{aligned}
&S_{100} + A_1 = 0 \\
&S_{200} + A_1 S_{100} + 2A_2 = 0 \\
&S_{300} + A_1 S_{200} + A_2 S_{300} + 3A_3 = 0 \\
&\quad \vdots \\
&S_{n00} + A_1 S_{(n-1)00} + \text{-----} + nA_n = 0
\end{aligned}
\tag{4}
$$

Now since the crystal is centrosymmetric,

$$
\begin{aligned}
A_n &= 1 \\
A_1 &= A_{n-1} \\
A_2 &= A_{n-2} \\
&\text{etc.}
\end{aligned}
\tag{5}
$$

By using equations (5) and eliminating the constants A_1, etc., in equations (4) Banerjee obtains a relationship among the values of S_h which must be satisfied by the observed values of the structure factor with the appropriate signs. Trial and error methods must then be used to determine these unknown signs. One must have more observed orders of hoo than parameters X_r in order to obtain a solution.

In order to apply this method to α-triglycine, the OOL zone reflections were employed. Allowing for the $P2_1/a$ space group extinctions, fourteen orders of OOL were observed on a well exposed Weissenberg film. Since there are 26 heavy atoms per cell with different x parameters there would be thirteen coefficients A_m to be eliminated in equation (4). Therefore in this case there were just enough equations to permit elimination of coefficients.

The problem was rendered more complicated by the fact

that only one of the fourteen 00L spectra was observed with zero intensity. Therefore the elimination of the A_n coefficients was made very tedious. Fortunately a method of writing down A_n as a function of the S_L values was found so that the calculations were accelerated somewhat.

A final equation relating the thirteen non-zero S_L quantities was finally derived. It contained ninety terms consisting of various cross-products of S_L and a constant term. Using the observed $|F_{00L}|$ on an absolute scale, various substitutions into this equation have been tried. Only one sign combination has been discovered so far which gives a residual within 5% of the constant term, but undoubtedly more exist.

It seems likely that similar equations for the 10L, 20L, 30L, etc. reflections will have to be derived before any real progress on this problem is achieved. Fourier line projections obtained from the 00L reflections with appropriate signs probably would not be well enough resolved to permit the accurate estimation of atomic parameters.

APPENDIX II.

PUNCHED CARD TECHNIQUES USED IN THIS INVESTIGATION

With the development of modern high-speed computing devices the entire field of multi-parameter organic crystal structures was made much more vulnerable to attack by ordinary X-ray methods. In particular, the use of punched card systems has made many of the computations involved in these complex structure determinations less cumbersome. Some of the applications of Hollerith or International Business Machine Corporation machines to various investigations have already been reported (41,42,43). Articles describing the summation of three dimensional Fourier series and least squares coefficients (44) and the calculation of structure factors (45) have also appeared. It is the purpose of this appendix to set forth some of the techniques utilized in the calculations for NN'-diglycyl-L-cystine which have not been previously published.

The computation of structure factors was performed using the International Business Machine's 604 electronic computing punch in conjunction with a 402 tabulator, a 513 reproducing punch, a sorting machine and a key punch. The methods employed were essentially the same as those outlined by Donohue and Schomaker for L-threonine (5). A and B for a general hkl reflection were computed from formulas 7b and 7c, the cosine and sine functions being obtained from series expansions of the arguments.

Usually a structure factor calculation for this compound was immediately followed by a least squares refinement of parameters. The necessary preliminary steps for this latter calculation, for example the obtaining of derivatives with respect to the coordinates and weighting factors, were made during the calculation of the structure factors themselves. The values of A and B for each reflection were summary punched to a set of detail cards, squared, added, and the resultant F^2_{calc} subtracted from the value of F^2_{obs} to give ΔF^2. It might be mentioned that the simultaneous calculation of structure factors and observational equation coefficients was materially aided by the wiring of a 604 plug board which could calculate the cosines and sines of two arguments in a single trip through the machine. From the nature of the equations 7b, 7c, and 17 it can be seen that such an arrangement would, in giving the cosine and sine of both $2\pi(hx_i + Lz_i)$ and $2\pi ky_i$ yields all the trigonometric functions needed for the two calculations.

The derivatives of all twelve atoms in the asymmetric unit for each observed non-zero reflection, together with the ΔF^2 term for that reflection, were multiplied by the appropriate weighting factor and transferred to a set of cards called observational equation cards. For the two and one dimensional least squares it was possible, allowing three digits for each coefficient, to put all the observational equation coefficients, the ΔF^2 term and an hkl identification into a single card. This was not possible for the thirty

seven terms in a typical three dimensional least squares observational equation, so three separate decks of cards representing the equations were prepared, one containing the derivatives with respect to x and z, another the derivatives with respect to y and x, and the last containing derivatives with respect to z and y.

Instead of employing the tabulational method described by Schaffer, Schomaker and Pauling (4), the 604 computing punch was used to obtain the normal equation coefficients. Three products Q_iQ_k, Q_iQ_l, and Q_iQ_m were computed from every card and added to the same products from other cards in suitable storage positions. Finally the $\sum_{hkl} Q_iQ_m$ sums were punched on blank trailer cards containing a characteristic x punch.

The actual machine time consumed in evaluating the coefficients of thirty six normal equations from about seven hundred observational equations was of the order of fifteen hours. This compares favorably with the three hours needed to obtain twelve normal equations from eighty observational equations using the tabulator as reported by Schaffer, Schomaker and Pauling (4). The time required to perform a calculation of structure factors for eight hundred hkl reflections from this structure was about nine hours, while the derivatives and ΔF^2 terms were computed and transferred to observational equation cards in five hours. Again these times are comparable with those cited by other authors for similar calculations.

VI. REFERENCES

1. G. Albrecht and R.B. Corey, J.A.C.S., 61, 1087 (1939).

2. H.A. Levy and R.B. Corey, J.A.C.S., 63, 2045 (1941).

3. D.P. Shoemaker, J. Donohue, V. Schomaker, and R.B. Corey, J.A.C.S., 72, 2328 (1950).

4. D.P. Shoemaker, R. Barieau, J. Donohue, and Chia-Si Lu, Acta Crystallographica, (to be published).

5. R.B. Corey, J.A.C.S., 60, 1598 (1938).

6. G.B. Carpenter and J. Donohue, J.A.C.S., 72, 2315 (1950).

7. A.B. Biswas and E.W. Hughes, Acta Crystallographica, (to be published).

8. E.W. Hughes and W.J. Moore, J.A.C.S., 71, 2618, (1949).

9. L. Pauling, "The Nature of the Chemical Bond", Cornell University Press (1945), p. 175.

10. E.J. Ambrose and A. Elliott, Pro. Roy. Soc., A205, 47 (1951).

11. R.B. Corey and J. Donohue, J.A.C.S., 72, 5349 (1950).

12. J.D. Bernal, Pro. Roy. Soc., A170, 75 (1939).

13. M.F. Perutz, Pro. Roy. Soc., A195, 474 (1949).

14. D. Harker, private communication (results as yet unpublished).

15. W.T. Astbury and H.J. Woods, Phil. Trans., 232, 333, (1933).

16. W.T. Astbury, C.E. Dalgliesh, S.E. Darmon and G.B.B.M. Sutherland, Nature, 162, 596 (1948).

17. M.L. Huggins, Chem. Rev., 32, 195 (1943).

18. W.T. Astbury and F.O. Bell, Nature, 147, 696 (1941).

19. W.L. Bragg, J.C. Kendrew, and M.F. Perutz, Pro. Roy. Soc., A203, 321 (1950).

20. E.J. Ambrose and W.E. Hanby, Nature, 163, 483 (1949).

21. C.H. Bamford, W.E. Hanby, and F. Happey, Pro. Roy. Soc., A205, 47 (1951).

22. L. Pauling and R.B. Corey, Pro. Nat. Acad. Sci., 37, 241 (1951).

23. M.F. Perutz, Nature, 167, 1053 (1951).

24. J.D. Bernal, Z. Krist., 78, 363 (1931).

25. F.V. Level, Z. Krist., 81, 224 (1932).

26. Chia-Si Lu, Rev. Sci. Inst., 14, 331 (1943).

27. A.J.C. Wilson, Nature, 150, 151 (1942).

28. J. Greenstein, J. Bio. Chem., 128, 241 (1939).

29. L. Pauling, "The Nature of the Chemical Bond", Cornell University Press (1945), p. 189.

30. Whitaker and Robinson, "The Calculus of Observations", Blackie and Son (1944), Chapter IX.

31. E.W. Hughes, J.A.C.S., 63, 1737 (1941).

32. D.W.J. Cruickshank, Acta Crystallographica, 2, 154 (1949).

33. A.J. Snow and R.E. Rundle, Acta Crystallographica, 4, 348 (1951).

34. J. Donohue, private communication.

35. J. Donohue, J.A.C.S., 72, 2701 (1950).

36. Margenau and Murphy, "The Mathematics of Physics and Chemistry", Van Nostrand, Inc., (1943), p. 486.

37. J. Vand, private communication.

38. W. Cochrane and F.H.C. Crick, Nature, 169, 234 (1952).

39. K. Banerjee, Pro. Roy. Soc., A141, 188 (1933).

40. E.W. Hughes, Acta Crystallographica, 2, 37 (1949).

41. A.D. Booth, "Fourier Technique in Organic Structure Analysis", Cambridge University Press (1948), p. 86.

42. J.M. Bennett and J.C. Kendrew, Acta Crystallographica, 5, 109 (1952).

43. J. Donohue and K.N. Trueblood, Acta Crystallographica, (to be published).

44. P.A. Shaffer, Jr., V. Schomaker, and L. Pauling, J. Chem. Phys., 14, 648 (1946).

45. J. Donohue and V. Schomaker, Acta Crystallographica, 2, 344, (1949).

VII. PROPOSITIONS

1. Many X-ray investigations of complex organic crystal structures have not made use of the phase determining Banerjee equations (1) because of the tedious nature of the elimination of coefficients from these relations. A method has been found, however, by which any of the coefficients a_n can be expressed in terms of the unitary structure factors F_h/f_h by inspection. If S_m^k represents the unit structure factor for plane m of the zone being studied raised to the positive integral power k, then

i) the expression for a_n must contain a sum of all terms $S_m^k S_l^j S_p^i$----- such that $km + jl + ip + \text{------} = n$;

ii) each term must be multiplied by a fraction equal to $1/m^k l^j p^i$-------k.'j ! i ! -------;

iii) the sign of each term is given by $(-1)^{k+j+i+\text{-------}}$.

The number of terms $S_m^k S_l^j S_p^i$---- for any given a_n is related to the number of terms for a_{n-2}, a_{n-4}, and a_{n-6} by a complicated set of difference equations.

2. J. Lewin (2) has reported the preparation of crystalline derivatives of several proteins with a number of metallic ions in which the mole ratio of metal to protein can be widely varied. If, as Lewin states, these ions are bound to the common acidic or basic functional groups of the protein molecule, it should be possible to prepare a series of such compounds with a single metallic complex and, by determining just the positions of the metal atoms, to obtain some idea of the relative positions of these functional groups

in the protein.

3. The weighting of least squares observational equations for refinement of structure parameters from X-ray data has been performed in a variety of ways. Weights dependent on the magnitudes of observed intensities, atom form factors, and independent unit weights have been used in different structure determinations. I propose that none of these systems adequately conforms to the nature of the errors encountered in visual intensity measurements since no simple dependence of these errors on the variables mentioned could be expected. The errors for each spectrum should be estimated individually from a consideration of all these factors. This procedure was used in the determination of the structure of L-threonine (3) but apparently has since been abandoned in favor of simpler approximations.

4. A differential equation relating the energy lost by an electron of initial velocity β per centimeter of its path through a body of density ρ has been given by Williams (4)

$$dV/dx = 1.06\rho\beta^{-1.4} \times 10^{6}$$

I propose that it is possible to integrate this equation, using series approximations, for several ranges of velocity β, and to use this analytic result in calculations involving the retardation of electrons in matter, in place of cumbersome graphical integration methods.

5. J.S. Ham and J.R. Platt (5) have stated that the absorption per peptide group at 1850A is about the same for

diglycine and triglycine, but that the lower absorption of diketopiperazine may be accounted for by a puckering of the six membered ring which would impair the planarity of the imide group. This explanation is not in accord with the results of the crystal structure determination of diketopiperazine in which Corey (6) found that the molecule is essentially planar. A more likely explanation would be based on either the fact that there are cis peptide links in diketopiperazine, as contrasted with the trans groups in di- and triglycine, or possible hyperconjugation effects in the ring system.

6. The diffuse nature of the X-ray reflections from fibers makes the quantitative estimation of their intensities difficult. N.C. Baenziger (7) has reported a technique whereby radioactive cobalt is deposited on negatives during the developing process in an amount proportional to the blackening of the film. I believe that this procedure might profitably be used to obtain quantitative intensity data from fiber photographs. The shapes of the reflections would make direct measurements on sections of the negatives impossible, but the radio-cobalt from each reflection area could be leached from the film and redeposited on uniform lead discs. Measurement of the activity of these discs would then give values of the integrated intensities on a relative scale.

7. In his book "Photosynthesis", Vol. I., Rabinowitch (8) discusses the structure of chlorophyll a in terms of three

structures which he considers isomeric. There seems to be no justification for supposing that the chlorophyll a structure allows such isomers to exist.

8. The methods available for the determination of molecular weights of peptide polymers often require the use of elaborate apparatus. It has been observed that solutions of poly-γ-methyl-L-glutamate in dimethyl-formamide show fluorescence effects which seem to vary with the degree of polymerization of the peptide. I propose that these effects be examined more closely since they may lead to a simple method of determining molecular weights and following the course of polymerization reactions.

9. In a recent paper to Nature, Bamford (9) claims that the calculated density of poly-γ-methyl-L-glutamate, using the 3.6 residue α-helix model, is too low to agree with the observed density of the fiber. I propose that this statement is erroneous; the parameters for this compound, as determined in this laboratory, give a calculated density of 1.398 gm/cm^3 for the 3.6 residue helix model. This value, in accordance with Bamford's own criteria, is in good agreement with the observed density.

10. It has been fairly well established that the polypeptide chain structure in poly-γ-methyl-L-glutamate has the 3.6 residue α-helix configuration, but the positions of the side chains in this molecule are still in doubt. (10) This may or may not be due to the complex character of the side chains

themselves. I propose that fibers of poly-serine or poly-valine be prepared and their diffraction patterns studied. Should these fibers possess an α phase, structure factor calculations might show whether or not the observed intensities are dependent on the parameters of their simple side chains.

11. J.H. Rayner and H.M. Powell (11) have shown that half the nickel atoms in crystalline $Ni(CN)_2NH_3C_6H_6$ are bonded to four cyanide carbon atoms in a planar configuration, and the other half form six octahedral bonds, four to cyanide nitrogens and two to ammonia groups. The coordination is thus very similar to that in $KFeFe(CN)_6$, but the magnetic moment of 2.32 Bohr magnetons is considerably greater than would be expected from the same analogy. I propose that similar palladium and platinum compounds (containing bi- and tetra-valent ions) be prepared and their magnetic moments determined in an effort to ascertain if this effect is common to the entire group of elements.

12. The various factors influencing the bond length in conjugated and aromatic molecules have been reviewed by C.A. Coulson (12). Bond order, the nature of the hybridization and the distribution of formal charges are all considered and the magnitude of their effects estimated. I propose that this treatment might be used to interpret some of the more complex bonds formed between metallic, semi-metallic, and non-metallic atoms.

REFERENCES FOR PROPOSITIONS

1. K. Banerjee, Pro. Roy. Soc., A141, 188 (1933).

2. J. Lewin, J.A.C.S., 73, 3906 (1951).

3. D.P. Shoemaker, J. Donohue, V. Schomaker, and R.B. Corey, J.A.C.S., 72, 2328 (1950).

4. E.J. Williams, Pro. Roy. Soc., A130, 310 (1932).

5. J.S. Ham and J.R. Platt, J. Chem. Phys., 20, 335 (1952).

6. R.B. Corey, J.A.C.S., 60, 1598 (1938).

7. N.C. Baenziger, U.S. Atomic Energy Commission AECD 2333 (1948).

8. E.I. Rabinowitch, "Photosynthesis", Vol. I., p. 442, Interscience Publishers, Inc., (1945).

9. C.H. Bamford, L. Brown, A.E. Elliott, W.E. Hanby, and I.F. Trotter, Nature, 169, 357 (1952).

10. H.L. Yakel, Thesis, California Institute of Technology, (1952).

11. J.H. Rayner and H.M. Powell, J. Chem. Soc., 1952, 319 (1952).

12. C.A. Coulson, J. Phys. Chem., 56, 311 (1952).